2.3.2. 3a·Toluene complex.

The 3a·toluene complex exhibits a packing arrangement that is essentially the same as that of 3a·benzene (Fig. 5). In both cases, the host framework of the complex is built by CH/π (edge-to-face) and 'bidentate' Ar–H/O= interactions. The guest toluene occupies the space between the host molecules by means of aromatic CH/π interactions of the two acenaphthene rings.
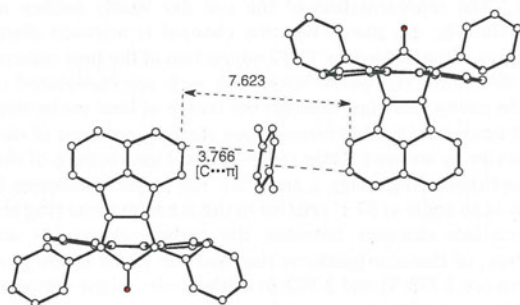

Fig. 5. Crystal packing of 3a·toluene complex (toluene presents a disorder).

The position of the guest methyl group is disordered in the crystal. The guest assumes two possible orientations related by 180° rotation around the centre of the benzene ring. On this basis, 3a might enclathrate p-xylene via a similar inclusion mode. The methyl group of toluene is extruded into a void space of the channel created by the host columns. The edge-to-face distance between the carbon atom of the acenaphthene ring and the plane of the guest π-system is 3.766(7) Å. The interatomic distance between the carbon atoms (3- and 4-position) of the acenaphthene ring via the guest molecule is approximately equal to that of 3a·benzene.

2.3.3. 3a·1,4-Dioxane complex.

The packing arrangements of 3a·1,4-dioxane is quite similar to that of the 3a·benzene complex (Fig. 6). The host molecules are held together by CH/π (edge-to-face) interactions to form a dimeric structure in which the 5-hydrogen atom of the acenaphthene ring is located at the opening of a V-shaped cavity created by the phenanthrene and acenaphthene ring, in which the closest distance between the carbon atom of the acenaphthene ring and the respective benzene ring of the phenanthrene plane is 3.763(5) Å. The hosts are also arranged in a head-to-tail fashion by the 'bidentate' CH/O-type hydrogen bonds [C···O, 3.413(5) and 3.422(4) Å] between the phenanthrene ring hydrogens and bridged carbonyl oxygen as well as the CH/π interactions between the offset-stacked acenaphthene moieties, thus forming a 'column'.

The guest dioxane is held between these stacking columns of the host molecules. Each ether oxygen of dioxane forms a CH/O-type hydrogen bond with the 2-hydrogen of the host phenyl group [C···O, 3.554(6) Å]. The analysis also suggests the presence of the CH/π interaction between the 2-methylene hydrogen of dioxane and the phenanthrene ring, in which the nearest C···C distance is 3.572(7) Å.

2.3.4. 3a·Acetone complex.

The crystal structure of the 3a·acetone complex has two independent hosts (hosts A and B) and an acetone molecule in the asymmetric unit. The network between host A molecules includes a 'bidentate' Ar–H/O interaction that is different from that observed in the 3a·benzene and 3a·1,4-dioxane complexes. The bridged carbonyl oxygen forms a CH/O-type hydrogen bond with the 4-hydrogen atom of the phenanthrene ring [O···C, 3.614(10) Å]. Another close contact is found between the carbonyl oxygen and the phenyl hydrogen of the adjacent host B [O···C, 3.279(10) Å].

On the other hand, the network pattern between host B molecules is also unique with regard to the relative disposition of the
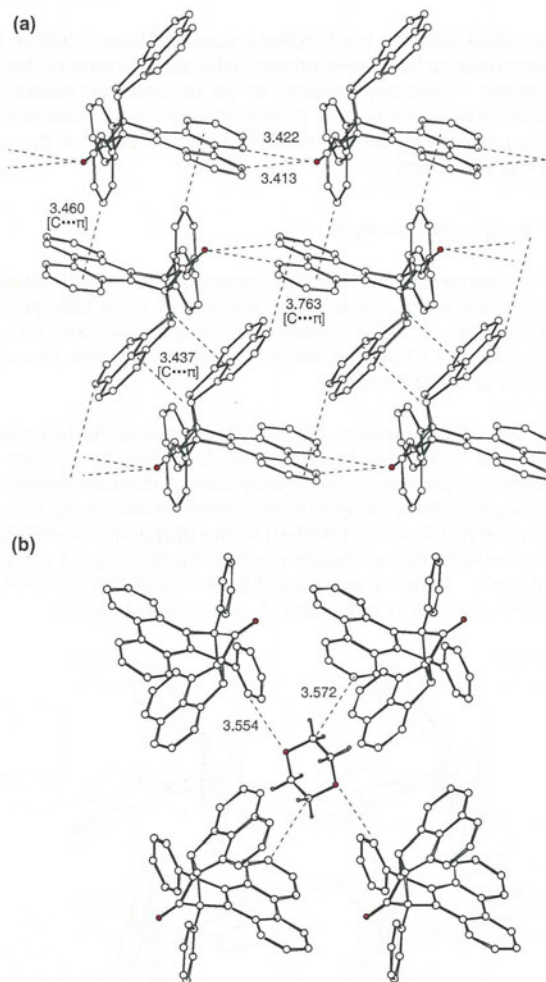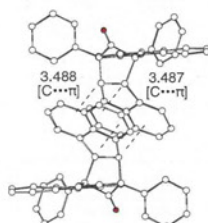

Fig. 6. Crystal packing of 3a·1,4-dioxane complex (a) host–host network (b) host–guest network.

host molecules. The carbonyl oxygen atom interacts with the 3- and 4-hydrogens of the phenanthrene ring, in which the C···O distances are 3.357(8) and 3.416(9) Å (Fig. 7).

The relative disposition of the host molecules affects the face-to-face stacking between neighbouring acenaphthene moieties. The distance between the methine carbon of the acenaphthene moiety and the neighbouring acenaphthene ring of the host A molecules is 3.483(7) Å [C–H···plane, 2.671(7) Å], whereas the corresponding separations are 3.488(7) and 3.487(7) Å [C–H···plane, 2.704(7) and 2.704(7) Å] with respect to host B molecules (Fig. 7b,c). Each type of host is independently linked head-to-tail to form a 'column' structure via 'bidentate' CH/O and CH/π interactions.

The phenanthrene rings of the host A are oriented in an edge-to-face manner with respect to host B. The 2-hydrogen atom of the phenanthrene ring is located perpendicular to the middle benzene ring of an adjacent phenanthrene ring with a C···plane distance of 3.876(9) Å (Fig. 8). In addition to the CH/O interaction between the carbonyl oxygen and the phenyl group, this interaction may play a leading role in the construction of the host framework.
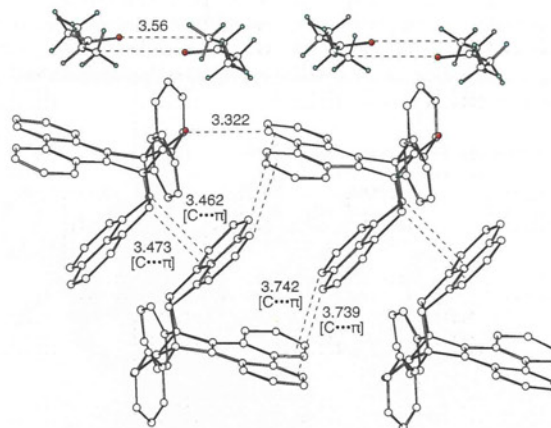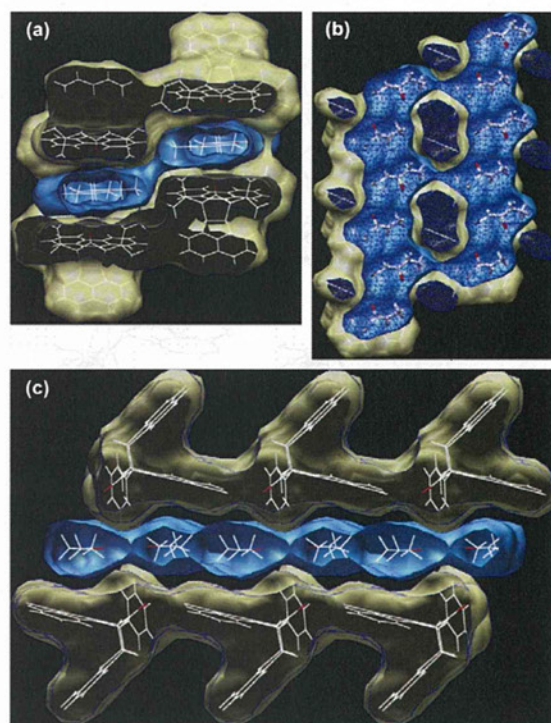
The guest acetone is located above the phenanthrene ring (region A) of host B and is accommodated in the space surrounded by two types of stacking host columns that are alternately arranged (Figs. 8 and 16b). Van der Waals forces are thought to be the primary interactions connecting the host and the guest in conjunction

(a) CH/O and CH/π interactions
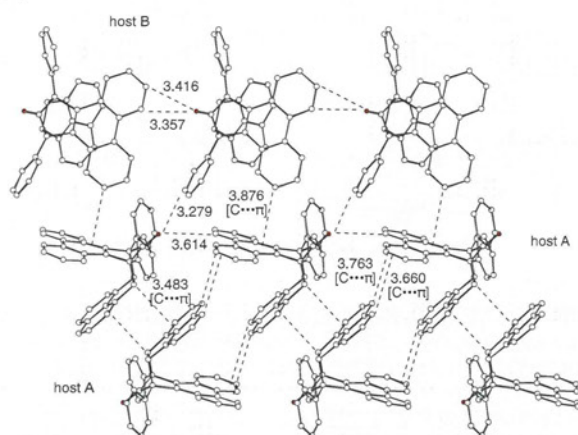
(b) CH/π interaction of host B **(c)** CH/π interaction of host A
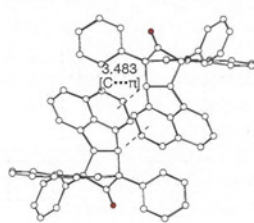
Fig. 7. Host–host network of **3a**·acetone complex.
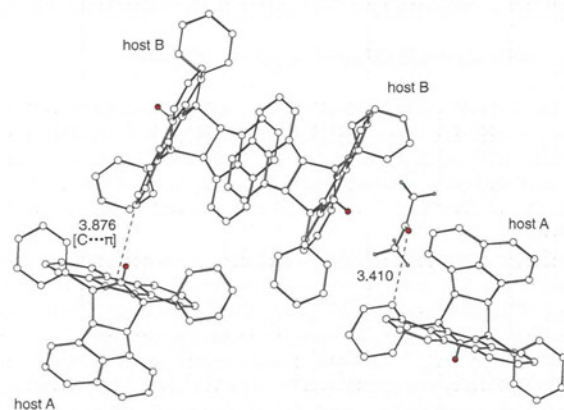
Fig. 8. Packing diagram of **3a**·acetone complex.

Fig. 9. Host–host and guest–guest network of **3a**·pentan-3-one complex.

Fig. 10. Packing diagram of **3a**·pentan-3-one complex (solid-surface representation).

with a weak Ar—H/O= interaction between the carbonyl oxygen of acetone and the hydrogen (2-position) of the phenyl group of host A [C···O, 3.410(13) Å].

*2.3.5. **3a**·Pentan-3-one complex.* The host–host network of the **3a**·pentan-3-one complex is stabilized by a 'bidentate' CH/O linkage and two types of CH/π interactions (i.e., Ar—H/π interactions between the acenaphthene hydrogen atoms and the phenanthrene ring and CH/π interactions between the methine hydrogens of the acenaphthene ring and the neighbouring acenaphthene ring). The distances are shown in Fig. 9.

The solid-surface representation of the clathrate indicates a characteristic layer-like structure having channels within a close distance along the crystallographic *c*-axis direction (Fig. 10). The

guest ketones are arranged along the host layer in such a way as to cancel the dipole–dipole interaction of the carbonyl functions. A pair of guest molecules forms a cyclic dimer unit by means of weak CH/O-type hydrogen bonds between the carbonyl oxygen and the hydrogen of the methyl group [C=O···(H)—C, 3.56(3) Å]. The guests form a layer-like structure located within the relatively large space created by the phenanthrene ring, bridged carbonyl group, and diphenyl group of the bicyclo[2.2.1]hept-2-en-7-one moiety, without remarkable attractive host-guest interactions. The edge-to-face (Ar—H/π) interactions between the phenanthrene ring plane and the phenyl ring of the adjacent host molecule found in **3a**·benzene and the Ar—H/π interactions between the phenanthrene rings found in **3a**·acetone are not observed in the **3a**·pentan-3-one complex, the absence of which should prove unfavourable for the stability of the host framework.

2.3.6. **3b**·*Acetone complex.* Two independent hosts (host A and B) are observed in the asymmetric unit of the crystal structure of the **3b**·acetone complex along with an acetone molecule. The packing diagram and notable host–host and host–guest interactions are shown in Figs. 11 and 12.
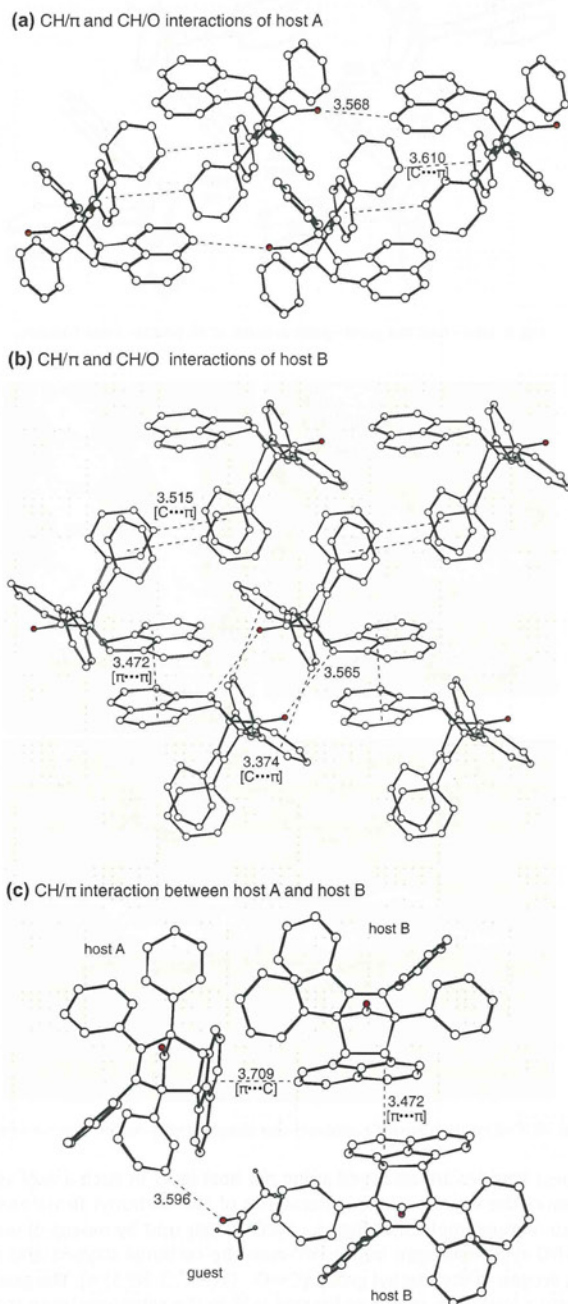
(a) CH/π and CH/O interactions of host A

(b) CH/π and CH/O interactions of host B

(c) CH/π interaction between host A and host B

Fig. 11. Host–host network of **3b**·acetone.

Both of the host molecules are loosely interconnected by the phenyl rings of the stilbene moieties to form a dimeric structure, in which the distances between the closest carbon atom (3-position) of a phenyl ring of the stilbene moiety and the phenyl ring plane of the neighbouring host molecule is 3.610(8) Å for host A and
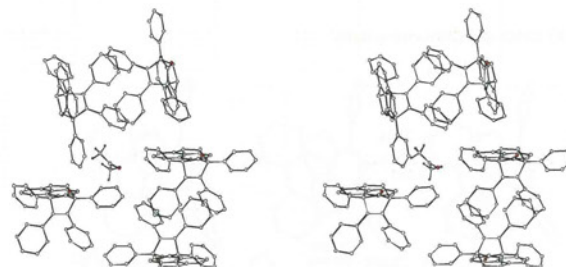

Fig. 12. Crystal packing of **3b**·acetone complex (stereoview).

3.515(7) Å for host B. The dimeric structures are connected infinitely along the *a*-axis by $>$C=O/H− hydrogen bonds between the carbonyl oxygen and the hydrogen of the acenaphthene ring of host A.

On the other hand, the dimer unit of host B is connected to the neighbouring unit by a CH/O-type hydrogen bond (between the carbonyl oxygen and methine carbon of the acenaphthene moiety) and an Ar−H/π interaction (between the phenyl rings). The interatomic and C···plane distances are 3.565(6) and 3.374(7) Å, respectively. Besides these interactions, the π/π interaction between the offset stacked acenaphthene rings of host B (plane···plane, 3.472 Å) and edge-to-face interaction between acenaphthene rings of hosts A and B [C···plane, 3.709(6) Å] stabilize the host-host network (Fig. 11c).

The guest acetone is accommodated in the space surrounded by two types of stacking host column structures, which are alternately arranged (Figs. 11 and 16d). The guest oxygen atom interacts with the hydrogen at the 3-position of a phenyl ring attached to the α-positions of the bridged carbonyl group by a $>$C=O/H−Ar interaction, and the methyl hydrogen of the guest acetone interacts with this phenyl ring by a CH/π interaction, respectively (Fig. 11c).

2.4. Molecular modelling of 3a·guest complex

We tried to assess the geometric features involving weak intermolecular interactions, such as edge-to-face or CH/O interaction forces and the energies of clathrate formation with guest molecules by semi-empirical MO calculations on the basis of successful application of the PM6[7−9] method to modelling of organic crystalline solids.[10,11]

The geometries of clathration models (involving one guest and eight hosts that interact with the guest as a minimal molecular unit) of the **3a**·benzene and **3a**·1,4-dioxane complexes were optimized by the PM6 method without imposing any symmetrical restrictions (Fig. 13).[12] However, the stationary point of **3a**·acetone (eight hosts and one guest) and **3a**·pentan-3-one complexes (eight hosts and four guests) could not be found by means of similar calculations. Based on the structural information of the clathration models, the stabilization energy due to the formation of the clathrate was estimated as the energy difference between ΔH_f (the clathration model) and ΔH_f (**3a**×8+guest) (Table 4).

The stabilization energy arising from intermolecular interactions of the host molecules in dimeric structures related by Ar−H/π (dimer A), CH/π (dimer B) and 'bidentate' CH/O (dimer C[13]) interactions (Fig. 14) was also calculated in a similar way. The energy of the various dimeric structures is 4–5 kcal/mol more stable than that of the components (Table 4). In the case of the host dimer formed by the 'bidentate' CH/O interaction, a CH/π interaction between the hydrogen of the phenanthrene ring and the phenyl ring of the neighbouring host was suggested. The degree of stabilization due to the purely CH/O interaction is assumed to be 2.5 kcal/mol on the basis of a calculation involving a model structure in which the two phenyl groups are replaced by hydrogen atoms.
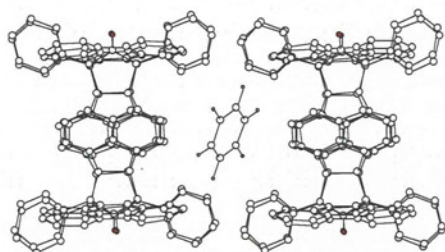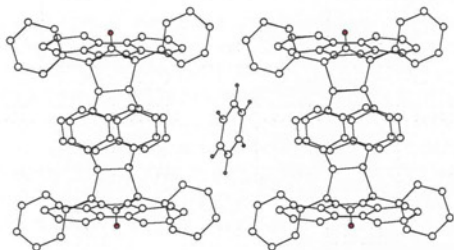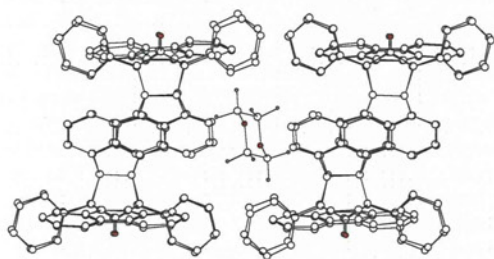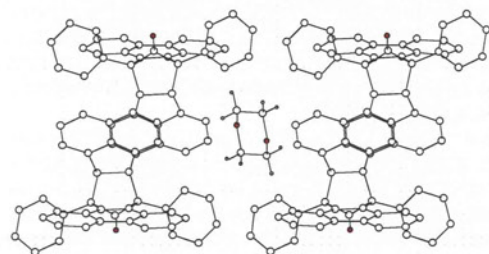
(a) PM6-optimized structure for the clathrate model of **3a**·benzene

(b) X-Ray structure of **3a**·benzene complex

(c) PM6-optimized structure for the clathrate model of **3a**·1,4-dioxane

(d) X-Ray structure of **3a**·1,4-dioxane complex

Fig. 13. PM6-calculated structure of **3a**·benzene complex and **3a**·1,4-dioxane (eight hosts and one guest) and the X-ray structures looking through the central guest molecule.

Table 4
PM6-optimized heats of formation (ΔH_f, kcal/mol) for the clathration model and the stabilization energies

Calculation model	ΔH_f	Stabilization energy, $\Delta\Delta H_f$[a]
Clathration model of **3a**·benzene (eight hosts and one guest)	1120.3	52.7
Clathration model of **3a**·1,4-dioxane (eight hosts and one guest)	1007.3	55.9
Host dimer by Ar–H/π interactions [dimer A]	283.2	4.0
Host dimer by CH/π interactions [dimer B]	282.0	5.2
Host dimer by 'bidentate' CH/O interactions including Ar–H/π interactions [dimer C]	283.5	3.7
Host (**3a**)	143.6	—
Benzene	24.2	—
1,4-Dioxane	−85.6	—

[a] $\Delta\Delta H_f = \Sigma\Delta H_f$ (components) − ΔH_f (calculation model).

(a) dimer A

3.597 [C···π]
2.668 [C-H···π]

(b) dimer B

3.418 [C···π]
2.492 [C-H···π]

(c) dimer C

3.689 [C···π]
3.229
3.230
3.692 [C···π]
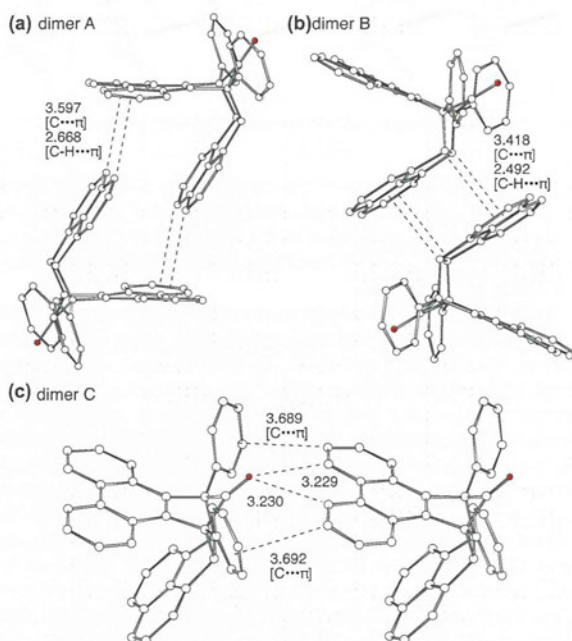
Fig. 14. PM6-calculated partial interaction structures: (a) host dimer by Ar–H/π (edge-to-face) interaction; (b) host dimer by CH/π interaction; (c) host dimer by 'bidentate' CH/O interaction.

3. Discussion

The structural feature of the host–host network of **3a**, in which the edge-to-face (between phenanthrene and phenyl rings) and 'bidentate' CH/O (between phenanthrene protons and carbonyl oxygen) interactions are operative, was consistent with those expected on the basis of the known clathrate of the carboxylic host.[3c] However, contrary to our expectations, the V-shaped cavity (region B) created by the phenanthrene and acenaphthene rings does not interact with aromatic guest molecules by aromatic CH/π or π/π interactions, but instead, the acenaphthene ring of an adjacent host molecule is included in this region (Fig. 15a). These additional aromatic CH/π interactions between the rigidly oriented

acenaphthene and phenanthrene rings necessarily confer additional strength to the host network, leading to the formation of stable clathrates. In fact, the desolvation temperatures of the **3a**·guest complexes are markedly higher than those of the corresponding **3b** complexes, the latter of which contains the conformationally flexible stilbene moiety. The stabilization of the host network of **3b** is significantly lowered by the lack of these remarkable attractive interactions between the aromatic rings of the host molecule. The host molecules of **3a** are also linked head-to-tail to form a 'column' structure by means of 'bidentate' CH/O hydrogen bonds between the phenanthrene ring hydrogens and bridged carbonyl oxygen as well as via CH/π interactions between the stacked acenaphthene moieties.

The crystal structures of the **3a**·guest complexes that were evaluated in this study show that the carbonyl oxygen atom contributes to the linkage of the host molecules without exception. This 'column' structure of the host/host network found in the clathrates of **3a** is a common characteristic feature (Fig. 15a). The cooperative effects of these weak intermolecular interactions

(a) Common 'column' structure

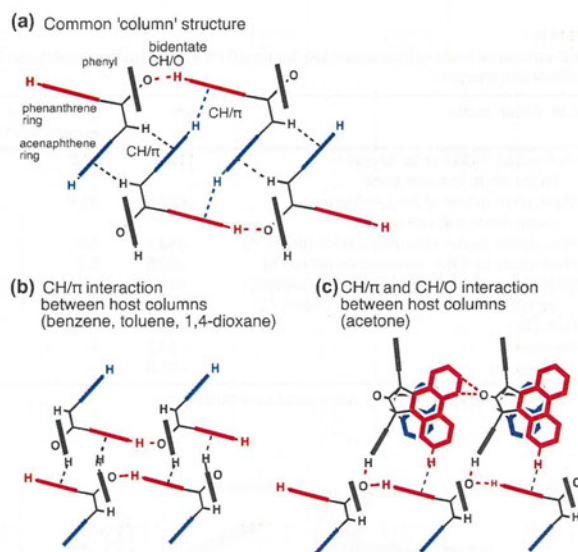**(b)** CH/π interaction between host columns (benzene, toluene, 1,4-dioxane)

(c) CH/π and CH/O interaction between host columns (acetone)

Fig. 15. Characteristic host/host network found in **3a**·guest complex.

should play a leading role in the construction of the host framework. The inability of **3a** to enclathrate methanol (see Table 3) can be attributed to the prevention of the formation of the head-to-tail linkage by typical hydrogen bonding between the bridged carbonyl of the host and methanol.

For the quantitative evaluation of weak interactions, such as CH/π, high-level quantum chemical calculations using MP2/cc-pVXZ (X=T or Q) or MP2/aug-cc-pVXZ (X=D or T) level, which include a large basis set and electron correlation correction are desirable in general.[14,15] However, the performance of such calculations for a crystal packing model consisting of a large number of molecules is still beyond the current computational feasibility. In order to compensate the disadvantage in the process of such high-level calculations, we examined the applicability of recently developed PM6 method that is aimed at providing a reasonable balance between speed and accuracy for investigation of molecular species of biochemical interest. The clathrates of **3a** seem to be a suitable model for testing applicability of PM6 to modelling of organic crystalline solids.

The spatial orientation of the guest molecule of the crystal structure as well as characteristic networks (Ar–H/π, CH/π, and 'bidentate' CH/O) between host molecules in the case of both clathrate models (**3a**·benzene and **3a**·1,4-dioxane) were approximately reproduced by the PM6 calculation (Fig. 13). For the clathration model of **3a**·benzene, the guest is held between the host molecules by aromatic CH/π (edge-to-face) interactions of the two acenaphthene rings, in which the edge-to-face distances between the carbon atoms (3- and 4-positions) of the acenaphthene ring and the plane of the guest π-system are 3.39 and 3.42 Å (the C–H···plane distances are 2.61 and 2.67 Å, respectively). In the clathrate model of **3a**·1,4-dioxane, CH/O interactions were found between the ether oxygen and the hydrogen of the phenyl group of the host. The closest distance between the carbon atom and the oxygen atom is 3.378 Å [X-ray, 3.554(6) Å]. Compared to the X-ray analysis data, these calculated distances seem to be somewhat underestimated. This may be due to the minimal clathration model, in which the interactions with the external environment of the model are oversimplified.

The stabilization energy of **3a**·benzene clathration model was estimated to be 52.7 kcal/mol, indicating that the formation of the clathrate generates a sizable stabilization. Similar stabilization was observed for the **3a**·1,4-dioxane clathration model (55.9 kcal/mol),

which has a packing structure similar to **3a**·benzene. The energy difference for **3a**·1,4-dioxane versus **3a**·benzene was ca. 3 kcal/mol, suggesting that enclathration of **3a** with 1,4-dioxane is thermodynamically more favourable than formation of the **3a**·benzene complex.

The stabilization energy of benzene arising from its inclusion in **3a** by means of double edge-to-face interactions is 4.0 kcal/mol, which was calculated from the difference between the optimized energy (ΔH_f) of the clathration model and the corresponding single point (1SCF) calculation in the absence of the guest. The value seems to be acceptable judging from the fact that the interaction energy of the T-shaped benzene dimer derived from CCSD(T) (coupled cluster calculations with single and double substitutions with non-iterative triple excitations) calculation is −2.46 kcal/mol at the maximum.[16]

The clathration model includes two dimer A, four dimer B and four dimer C units (see Fig. 14). The sum of the stabilization energies from each of these units is 43.6 kcal/mol, which accounts for a major portion (ca. 83%) of the stabilization energy derived by formation of the clathrate, suggesting that the weak Ar–H/π, CH/π and 'bidentate' CH/O interactions associated with dimer formation play a major role in the construction of the host framework. PM6 method seems to be a useful tool for estimation of interaction in organic crystalline solids that are hard to be subjected to high-level MO calculation.

The additional CH/π interactions between the phenyl ring and the phenanthrene ring plane of the adjacent host molecule (interactions between the column structures) in the **3a**·benzene complex further strengthen the host/host network (Fig. 15b). A similar structural feature is found in the **3a**·toluene and **3a**·1,4-dioxane complexes, which exhibit high desolvation temperatures in comparison to the boiling point of the pure guests. In the case of the **3a**·acetone complex, the host framework is constructed by CH/O interactions between the bridged carbonyl oxygen and the phenyl ring hydrogen as well as the CH/π interaction between the phenanthrene rings in lieu of the CH/π interactions found in **3a**·benzene (Fig. 15c), whereas such interactions between the hosts are inoperative in the case of **3a**·pentan-3-one.

The flexible network constructed by the weak host–host interactions is guest-inclusion dependent whereby the arrangement of the columns is affected by the size and character of the guest molecule; the guest molecules might act as 'glue' that binds the host column structures. The inclusion pattern of **3a** clarified in this study is shown in Fig. 16a–c.

The host molecule (**3a**) enclathrated ketones, cyclic ethers and aromatics. Benzene was particularly strongly incorporated within the host lattice, and the desolvation temperatures were extremely high relative to the boiling point of the pure guest (Table 2). The host lattice contains channels along the crystallographic c-direction. The aromatic guests are located in the spaces between region C of the host columns and are held by edge-to-face (CH/π) interactions between acenaphthene ring hydrogens and the benzene ring (Fig. 16a). The importance of the rigid acenaphthene moiety for aromatic guest recognition is supported by the fact that the endo [4+2]π cycloadduct of **1a** and styrene did not show inclusion ability.[3a] On the other hand, the desolvation temperatures of toluene and p-xylene enclathrated in **3a** were not remarkably high in comparison with the boiling point of the pure liquid, suggesting that steric crowding within the space weakened host lattice. However, the possibility that the packing pattern of **3a**·m-xylene is different from that of **3a**·benzene cannot be denied because the desolvation temperature is exceptionally low in the latter complexes. Judging from the cavity of **3a**·toluene, sufficient space may not be available in the void to extrude the m-methyl group of m-xylene.

Ketones tend to occupy the space within region A. In the **3a**·acetone complex, the host network is strengthened by an edge-
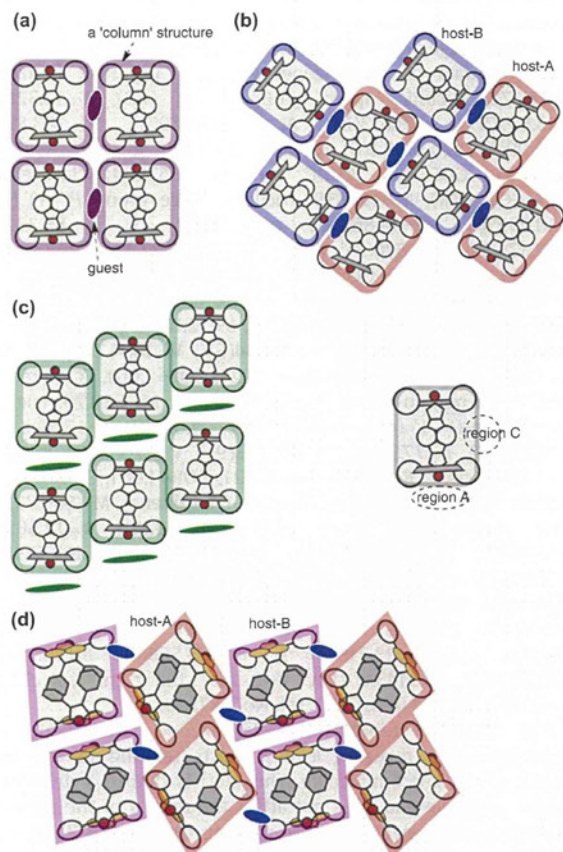
Fig. 16. Schematic drawing of inclusion mode of host 3a: (a) 3a·benzene or 3a·1,4-dioxane (2:1). (b) 3a·Acetone (2:1). (c) 3a·Pentan-3-one (1:1). (d) 3b·Acetone (2:1).

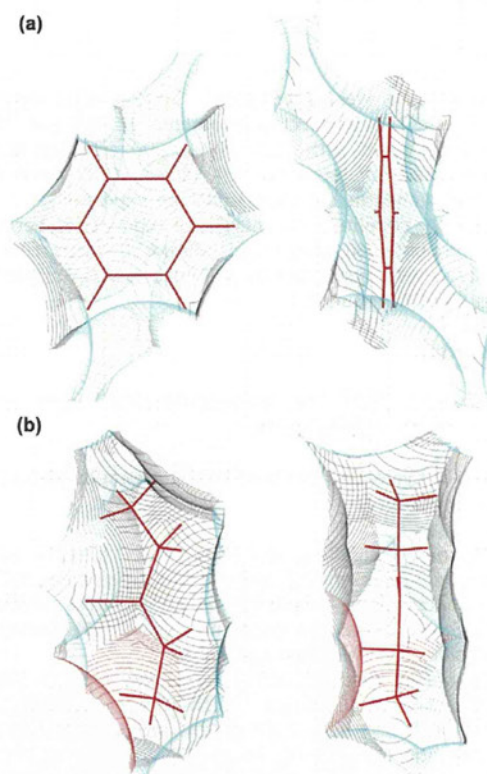Fig. 17. Composite diagram of the cavity (a) 3a·benzene (b) 3a·pentan-3-one (The radius of each sphere is taken to be 1.2 Å greater than the van der Waals radius of the corresponding atom.).

to-face interaction between the phenanthrene rings of adjacent hosts (Fig. 15c). The difference between the onset temperature for desolvation and the boiling point of the pure solvent for 3a·pentan-3-one is much lower than that for 3a·acetone, which may be explained by the difference in their inclusion modes as well as the strength of the host—host network stated above.

Fig. 17 shows the cavities limited by a concave surface of the sphere from the surrounding atoms, which was created by using the program CAVITY.[17] The cavity shows the area accessed by the centres of atoms of the guest. The guest 3a·pentan-3-one is loosely fitted in the cavity encapsulated by the host and guest molecules, whereas 3a·benzene exhibits the typical tight fitting in which there are several significant close contacts between the carbon atoms of the host and guest. Only CH/O-type interactions between the guest molecules are found in 3a·pentan-3-one; CH/π-type close contacts between the host and the guest molecules also confer a stabilizing effect on the crystal packing. The dimeric units of 3a·pentan-3-one also form a layer that is interdigitated with region A of the host columns, without remarkable attractive interactions between the dimeric units. The absence of these interactions in conjunction with the void space might facilitate the ready collapse of the clathrate. The inclusion behaviour exhibited by host 3b, which contains the stilbene moiety, is different from that of host 3a in that the aromatic guests are included in a 1:1 host/guest ratio. The desolvation temperatures of the 3b clathrates are appreciably lower than those observed with host 3a. The replacement of the phenanthrene moiety of 3a with the stilbene moiety in 3b may decrease the stability of the 3b host—host network due to the absence of both

the 'bidentate' CH/O interaction between the phenanthrene ring and the bridged carbonyl oxygen as well as the Ar—H/π interaction between phenanthrene and the phenyl rings. As pointed out above, the host/host network of 3b is primarily constructed by interactions among the phenyl rings. These interactions among the conformationally flexible phenyl groups are considered to be relatively weak as reflected in the DTA data.

On the other hand, the acenaphthene ring plays an important role in the construction of the host—host network of 3b. CH/π and π/π interactions between the acenaphthene rings are operative in 3b·acetone (Fig. 11). This shows that a suitable spatial arrangement of the aromatic recognition motif (i.e., acenaphthene, phenanthrene and phenyl rings) will produce similar inclusion behaviour in various host molecules. This information should prove useful in the design of suitable hosts for recognition of aromatic guests via CH/π and π/π interactions.

4. Conclusion

This study demonstrated that the clathrate hosts (3a) derived from the [4+2]π cycloaddition of phencyclone (1a) to acenaphthylene (2) effectively furnishes the aromatic CH/π (edge-to-face) interactions between the phenanthrene unit and acenaphthene ring, leading to strong host—host networks. These weak interactions as well as 'bidentate' CH/O interactions between phenanthrene-ring hydrogens and a bridged carbonyl group contribute to the formation of characteristic host 'column' structures, which facilitates the inclusion of various guest molecules within the void space between the host columns.

5. Experimental section

5.1. General

Melting points were uncorrected. The IR spectra were taken with a Hitachi 270–30 spectrophotometer. ^{1}H NMR and ^{13}C NMR spectra were taken with JEOL JNM-A 500 and JNM-ECA 500 (500 MHz) spectrometers for ca. 10% solution (CDCl$_3$) with TMS as an internal standard; chemical shifts are expressed as δ values (ppm) and the coupling constants (J) are expressed in hertz. Mass spectra were obtained using a JEOL JMS-DX 303 instrument. Thermal analyses were performed on a Shimadzu DTG-50/50H Simultaneous TG/DTA instrument.

5.2. Materials

Phencyclone (**1a**)[18] and tetracyclone (**1b**)[19] were prepared according to the reported methods.

5.3. [4+2]π Cycloadducts (3) of 1 with acenaphthylene (2) (general procedure)

A solution of **1a** (3.82 g, 10 mmol) and **2** (1.83 g, 12 mmol) in toluene (20 mL) was refluxed until the dark green colour had faded out (2 h). After cooling, the precipitated crystals were collected and washed with cold ether and dried under vacuum. Recrystallization from benzene gave colourless needles **3a**.[20]

Compound **3a**: Yield 83%, colourless needles; mp 307–308 °C. IR (KBr): 1782 (bridged $>$C=O) cm^{-1}. ^{1}H NMR (500 MHz, CDCl$_3$) δ: 5.38 (s, 2H, *exo*-methine), 6.98–7.80 (m, 20H, aromatic H), 8.28 (d, 2H, J=7.9, aromatic H), 8.31 (d, 2H, J=7.3, aromatic H). ^{13}C NMR (125 MHz, CDCl$_3$) δ: 48.5 (methine), 66.2 ($>$C$<$), 120.9, 122.6, 123.2, 125.2, 125.6, 126.2, 127.1, 127.6, 128.0, 129.0, 129.1, 132.1 (aromatic CH), 126.9, 130.4, 130.5, 134.6, 135.9, 141.8, 142.3 (aromatic $>$C$<$), 201.8 (bridged $>$C=O). MS (EI, m/z): 534 (M$^+$), 506 (M$^+$–CO). Anal. Calcd for C$_{41}$H$_{26}$O: C, 92.11; H, 4.90. Found: C, 91.82; H, 5.13.

Compound **3b**:[21] Yield 78%, colourless prisms; mp 215–217 °C. IR (KBr): 1774 (bridged $>$C=O) cm^{-1}. ^{1}H NMR (500 MHz, CDCl$_3$) δ: 5.16 (s, 2H, *exo*-methine), 6.06 (d, 4H, J=7.9, aromatic H), 6.61–6.64 (m, 4H, aromatic H), 6.75–6.78 (m, 2H, aromatic H), 7.14–7.65 (m, 16H, aromatic H). ^{13}C NMR (125 MHz, CDCl$_3$) δ: 49.2 (methine), 68.1 ($>$C$<$), 122.2, 123.9, 126.4, 126.9, 127.5, 127.7, 128.2, 129.1, 130.5 (aromatic CH), 131.7, 134.4, 134.5, 142.2, 142.7, 143.4 (aromatic $>$C$<$), 200.7 (bridged $>$C=O). MS (EI, m/z): 536 (M$^+$), 508 (M$^+$–CO). Anal. Calcd for C$_{41}$H$_{28}$O: C, 91.76; H, 5.26. Found: C, 92.00; H, 5.31.

5.4. Preparation of inclusion complexes

The host compound was dissolved under heating in the minimum amount of a guest solvent. The solution was allowed to stand for several days at room temperature. The crystals were collected by suction filtration and dried.

5.5. X-ray crystallography

The single crystals for inclusion complexes were prepared by slow evaporation of solution at room temperature. All measurements were performed on a Rigaku RAXIS RAPID imaging plate area detector with graphite-monochromated Mo-Kα radiation (λ=0.7107 Å). The data were collected at a temperature of 23±1 °C to a maximum 2θ value of 55°. The structure was solved by direct method (SIR2008[22a]), and all hydrogen atoms were located at calculated positions. The structure was refined by a full-matrix least-squares technique using anisotropic thermal parameters for non-hydrogen atoms and a riding model for hydrogen atoms. All calculations were performed using the crystallographic software package Crystal Structure.[22b,c] These X-ray crystallographic data have been deposited at the Cambridge Crystallographic Data Centre (CCDC).

Crystal data of **3a**·benzene: C$_{41}$H$_{26}$O·C$_3$H$_3$ (1/2 benzene), M=573.71, triclinic, space group $P(-1)$ (#2), a=12.328(2), b=13.637(5), c=9.613(2) Å, α=103.89(2), β=103.83(1), γ=95.09(2)°, V=1505.1(6) Å^3, Z=2, D_c=1.266 g cm^{-3}, R=0.069, R_w=0.081, GOF=1.16, CCDC ref. No. 857594; **3a**·acetone: 2C$_{41}$H$_{26}$O·C$_3$H$_6$O, M=1127.39, triclinic, space group $P(-1)$ (#2), a=10.0359(8), b=14.863(2), c=20.408(2) Å, α=88.332(3), β=87.334(2), γ=79.546(3)°, V=2989.7(5) Å^3, Z=2, D_c=1.252 g cm^{-3}, R=0.1271, R_w=0.1870, GOF=1.020, CCDC ref. No. 857595; **3a**·1,4-dioxane: C$_{41}$H$_{26}$O·C$_2$H$_4$O (1/2 1,4-dioxane), M=578.71, triclinic, space group $P(-1)$ (#2), a=9.679(1), b=12.103(2), c=13.501(2) Å, α=86.521(6), β=76.752(4), γ=76.084(5)°, V=1494.2(4) Å^3, Z=2, D_c=1.286 g cm^{-3}, R=0.0789, R_w=0.1224, GOF=1.001, CCDC ref. No. 857596; **3a**·pentan-3-one: C$_{41}$H$_{26}$O·C$_5$H$_{10}$O, M=620.79, triclinic, space group $P(-1)$ (#2), a=9.849(3), b=11.896(3), c=16.596(5) Å, α=70.257(8), β=77.572(9), γ=69.623(9)°, V=1705.2(8) Å^3, Z=2, D_c=1.209 g cm^{-3}, R=0.1107, R_w=0.1821, GOF=1.001, CCDC ref. No. 857597; **3a**·toluene: C$_{41}$H$_{26}$O·C$_{3.5}$H$_4$ (1/2 toluene), M=580.73, triclinic, space group $P(-1)$ (#2), a=9.5582(9), b=12.509(2), c=13.767(2) Å, α=95.406(4), β=104.253(3), γ=103.573(4)°, V=1530.6(3) Å^3, Z=2, D_c=1.260 g cm^{-3}, R=0.0742, R_w=0.1067, GOF=1.167, CCDC ref. No. 857598; **3b**·acetone: 2C$_{41}$H$_{28}$O·C$_3$H$_6$O, M=1131.42, triclinic, space group $P(-1)$ (#2), a=11.0417(6), b=12.351 (1), c=23.719(2) Å, α=103.558(3), β=90.791(3), γ=90.867(4)°, V=3143.8(4) Å^3, Z=2, D_c=1.195 g cm^{-3}, R=0.1018, R_w=0.1723, GOF=1.000, CCDC ref. No. 857599.

The checkCIF reports for **3a**·benzene, **3a**·toluene and **3a**·pentan-3-one revealed level-A alerts relating to the bond lengths of the guest molecule. This alert is due to a disorder of a guest molecule and does not affect the results of the structure determination. Inspection of the packing structure of **3b**·acetone suggests that the guest acetone assumes several possible orientations related by rotation around the carbonyl carbon atom; this has resulted in a level-A alert in the CIF check.

Acknowledgements
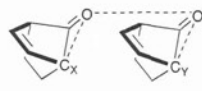
Supplementary data

Further crystal packing drawings and DTA/TG diagrams. Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.tet.2012.03.005.

References and notes

1. (a) MacNicol, D. D., Toda, F., Bishop, R., Eds. ; Pergamon: Oxford, 1996; Vol. 6; (b) Weber, E. In *Inclusion Compounds*; Atwood, J. L., Davies, J. E. D., MacNicol, D. D., Eds.; Oxford University: Oxford, 1991, Chapter 5; (c) Steed, J. W.; Atwood, J. L. *Supramolecular Chemistry*; Wiley: Chichester, UK, 2000.
2. (a) Weber, E.; Csoregh, I.; Ahrendt, J.; Finge, S.; Czugler, M. *J. Org. Chem.* **1988**, *53*, 5831–5839; (b) Weber, E.; Skobridis, K.; Wierig, A.; Nassimbeni, L. R.; Johnson, L. *J. Chem. Soc., Perkin Trans. 2* **1992**, 2123–2130; (c) Ochiai, K.; Mazaki, Y.; Nishikiori, S.; Kobayashi, K.; Hayashi, S. *J. Chem. Soc., Perkin Trans. 2* **1996**, 1139–1145; (d) Ibragimov, B.; Beketov, K.; Makhkamov, K.; Weber, E. *J. Chem. Soc., Perkin Trans. 2* **1997**, 1349–1352; (e) Weber, E.; Hens, T.; Li, Q.; Mak, T. C. W. *Eur. J. Org. Chem.* **1999**, 1115–1125; (f) Kishikawa, K.; Tsubokura, S.; Kohmoto, S.; Yamamoto, M. *J. Org. Chem.* **1999**, *64*, 7568–7578; (g) Weber, E.; Hens, T.; Brehmer, T.; Csoregh, I. *J. Chem. Soc., Perkin Trans. 2* **2000**, 235–241; (h) Caira, M. R.; Nassimbeni, L. R.; Vujovic, D.; Weber, E. *J. Chem. Soc., Perkin Trans. 2* **2001**, 861–863; (i) Weber, E.; Nitsche, S.; Wierig, A.; Csoregh, I. *Eur. J. Org. Chem.* **2002**, 856–872; (j) Caira, M. R.; Roex, T.; Nassimbeni, L. R.; Ripmeester, J. A.; Weber, E. *Org. Biomol. Chem.* **2004**, *2*, 2299–2304; (k) Bourne, S. A.; Corin, K.; Cruickshank, D. L.; Davson, J.; Nassimbeni, L. R.; Su, H.; Weber, E. *New J. Chem.* **2011**, *35*, 1556–1561.

3. (a) Eto, M.; Setoguchi, K.; Harada, A.; Sugiyama, E.; Harano, K. *Tetrahedron Lett.* **1998**, *39*, 9751–9754; (b) Yoshitake, Y.; Misaka, J.; Setoguchi, K.; Abe, M.; Kawaji, T.; Eto, M.; Harano, K. *J. Chem. Soc., Perkin Trans. 2* **2002**, 1611–1619; (c) Yoshitake, Y.; Misaka, J.; Abe, M.; Yamasaki, M.; Eto, M.; Harano, K. *Org. Biomol. Chem.* **2003**, *1*, 1240–1249.

4. (a) Taylor, R.; Kennard, O. *J. Am. Chem. Soc.* **1982**, *104*, 5063–5070; (b) Desiraju, G. R. *Acc. Chem. Res.* **1991**, *24*, 290–296; (c) Desiraju, G. R. *Acc. Chem. Res.* **1996**, *29*, 441–449; (d) Steiner, T.; Kanters, J. A.; Kroon, J. *Chem. Commun.* **1996**, 1277–1278; (e) Steiner, T. *Crystallogr. Rev.* **1996**, *6*, 1–51; (f) Steiner, T. *Angew. Chem., Int. Ed.* **2002**, *41*, 48–76.

5. (a) Nishio, M.; Hirota, M.; Umezawa, Y. *The CH/π Interaction; Evidence Nature and Consequences*; Wiley: New York, NY, 1998; (b) Takahashi, O.; Kohno, Y.; Nishio, M. *Chem. Rev.* **2010**, *110*, 6049–6076; (c) Nishio, M. *Phys. Chem. Chem. Phys.* **2011**, *13*, 13873–13900; (d) Suezawa, H.; Yoshida, T.; Hirota, M.; Takahashi, H.; Umezawa, Y.; Honda, K.; Tsuboyama, S.; Nishio, M. *J. Chem. Soc., Perkin Trans. 2* **2001**, 2053–2058; (e) Recent lists of articles relating CH/π interaction are available from http://www.tim.hi-ho.ne.jp/dionisio/page/other.html.

6. The figures are generated by UCSF Chimera program. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, *25*, 1605–1612.

7. Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209–220.

8. MOPAC2009 Version 9.062M: Stewart, J. J. P., Stewart Computational Chemistry (http://OpenMOPAC.net).

9. N. Senda, Idemitsu Technical Report, 2006, *49*, pp 106–111.

10. Stewart, J. J. P. *J. Mol. Model.* **2008**, *14*, 499–535.

11. Eto, M.; Yamaguchi, K.; Shinohara, I.; Ito, F.; Yoshitake, Y.; Harano, K. *Tetrahedron* **2011**, *67*, 7400–7405.

12. Energy minimization of the structures was performed using EF (eigenvector following routine) algorithm, with a gradient norm (GNORM) value set to 0.01. The calculation data (atomic coordinates) are available upon request (e-mail address: meto@ph.sojo-u.ac.jp).

13. The dihedral angle of C_X–O–O–C_Y between two host molecules was fixed to 0° according to the X-ray structure.

14. Shibasaki, K.; Fujii, A.; Mikami, N.; Tsuzuki, S. *J. Phys. Chem. A* **2006**, *110*, 4397–4404.

15. A reviewer pointed out that the analysis of weak noncovalent interactions with energy by semi-empirical MO calculation seems not to be good.

16. Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. *J. Am. Chem. Soc.* **2002**, *124*, 104–112.

17. Ohashi, Y.; Yanagi, K.; Kurihara, T.; Sasada, Y.; Ohgo, Y. *J. Am. Chem. Soc.* **1982**, *104*, 6353–6359.

18. Forsyth, W. R.; Weisenburger, G. A.; Field, K. W. *Trans. Illinois State Acad. Sci.* **1996**, *89*, 37–40.

19. Allen, C. F. H. *Chem. Rev.* **1945**, *37*, 209–268.

20. Sasaki, T.; Kanematsu, K.; Iizuka, K. *J. Org. Chem.* **1976**, *41*, 1105–1112.

21. Mondal, S.; Bandyopadhyay, T. K.; Bhattacharya, A. *J. Indian J. Chem., Sect. B: Org. Chem. Incl. Med. Chem.* **1983**, *22*, 225–229.

22. (a) SIR2008: Burla, M. C.; Caliandro, R.; Camalli, M.; Carrozzini, B.; Cascarano, G.; De Caro, L.; Giacovazzo, C.; Polidori, G.; Siliqi, D.; Spagna, R. *J. Appl. Crystallogr.* **2007**, *40*, 609–613; (b) CrystalStructure 4.0: Crystal Structure Analysis Package; Rigaku Corporation: Tokyo 196-8666, Japan, 2000–2010; (c) Carruthers, J. R.; Rollett, J. S.; Betteridge, P. W.; Kinna, D.; Pearce, L.; Larsen, A.; Gabe, E. *CRYSTALS Issue 11*; Chemical Crystallography Laboratory: Oxford, UK, 1999.

RETROVIROLOGY

REVIEW

Open Access

Host restriction factors in retroviral infection: promises in virus-host interaction

Yong-Hui Zheng[1*], Kuan-Teh Jeang[2] and Kenzo Tokunaga[3]

Abstract

Retroviruses have an intricate life cycle. There is much to be learned from studying retrovirus-host interactions. Among retroviruses, the primate lentiviruses have one of the more complex genome structures with three categories of viral genes: structural, regulatory, and accessory genes. Over time, we have gained increasing understanding of the lentivirus life cycle from studying host factors that support virus replication. Similarly, studies on host restriction factors that inhibit viral replication have also made significant contributions to our knowledge. Here, we review recent progress on the rapidly growing field of restriction factors, focusing on the antiretroviral activities of APOBEC3G, TRIM5, tetherin, SAMHD1, MOV10, and cellular microRNAs (miRNAs), and the counter-activities of Vif, Vpu, Vpr, Vpx, and Nef.

Review

As an obligatory intracellular parasite with limited genome size, retroviruses interact with both supportive and inhibitory host factors to complete their life cycle. Supportive factors help the virus enter the cell, duplicate its viral genome, make viral proteins, and release new progeny particles, while inhibitory factors could, in principle, intervene against the virus at every step of replication. These inhibitory factors are collectively called host restriction factors. Unique from other viruses, retroviruses require replication steps such as RNA reverse transcription and DNA integration, which create additional targets for restriction. Historically, the first host restriction for retroviruses was discovered when murine leukemia virus (MLV) infection was found to be inhibited by the Friend virus susceptibility factor-1 (Fv1) [1]. In the mouse genome, there are at least two $Fv1$ alleles ($Fv1^n$, $Fv1^b$) that confer resistance to B-tropic MLV (B-MLV) or N-tropic MLV (N-MLV) infection. The B-MLV strains efficiently infect $Fv1^{b/b}$ homozygous Balb/c mice but not the $Fv1^{n/n}$ homozygous NIH/Swiss mice, whereas the N-MLV strains have an opposite tropism. The $Fv1$ gene is located on mouse chromosome 4 [2], which encodes an endogenous retrovirus Gag-like protein [3]. Fv1 recognizes the MLV capsid (CA) protein

through a single residue at position 110 [4], and blocks the nuclear import of reverse transcribed retroviral pre-integration complex [5], but the precise mechanism is still unknown.

The initial observations on Fv1 have been followed by the discovery of additional restriction factors in mammalian cells [6]. In general, restriction factors have been identified from non-permissive cells, where virus replication is severely restricted. The restriction phenotype can be dominant and potent and can suppress viral replication up to several orders of magnitude. Thus, when a non-permissive cell is fused with a permissive cell, the heterokaryon inherits the restrictive phenotype. Another finding with restriction factors is that they can display hallmarks of positive genetic selection during evolution, indicative of their beneficial advantage to the host in settings of host-pathogen conflicts. Moreover, restriction factors can be constitutively or inducibly (e.g. induced by interferon) expressed, and many viruses have evolved countervailing stratagems to neutralize the activities of restriction factors. For example, HIV-1 co-opts the host ubiquitin/proteasome system (UPS) to degrade cellular restriction factors. The recent advances accrued from studying restriction factors have expanded our views on virus-host interaction as well as host innate immunity to viral infections [7]. The findings have not only provided a more comprehensive understanding of the virus life cycle, but have also offered clues on new antiviral mechanisms and targets. Below, we survey the antiviral

* Correspondence: zhengyo@msu.edu
[1]Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI, USA
Full list of author information is available at the end of the article

activities of several cellular restriction factors that impede HIV replication, including APOBEC3, TRIM5α, tetherin, SAMHD1, MOV10 and cellular miRNAs; and we outline viral countermeasures to subdue these restrictions (Figure 1).

APOBEC3 proteins

APOBEC3 (A3) refers to <u>A</u>polipoprotein <u>B</u> mRNA-editing enzyme <u>c</u>atalytic polypeptide-like <u>3</u> proteins, which include A3A, A3B, A3C, A3DE, A3F, A3G, and A3H. These genes are tandemly arrayed on human chromosome 22 and were discovered by homology search using the APOBEC1 signature sequence [8,9]. The *A3DE* gene was initially proposed as two separated genes (*A3D, A3E*); later, it was shown to produce alternatively spliced mRNAs from a single gene. Unlike humans, the mouse genome encodes only a single *A3*

gene. Altogether these proteins belong to the cytidine deaminase family that includes four additional members: activation-induced cytidine deaminase (AID), APOBEC1, APOBEC2, and APOBEC4. All APOBEC proteins have one or two copies of zinc-coordinating deaminase domain (the Z domain) with a signature motif $HxEx_{23-28}PCxxC$, which catalyzes cytidine deamination that converts cytosines to uracils (C to U) on a variety of vertebrate-specific RNA/DNA targets (Figure 2). APOBEC1 regulates lipid metabolism [10], and AID contributes to antibody production [11]. The physiological function of APOBEC2 and APOBEC4 is still unknown. A3A, A3C, and A3H have one copy of Z domain, whereas A3B, A3DE, A3F, and A3G have two. The functions of these A3 proteins on viral infection were not clear, until it was discovered that A3G has a very potent anti-HIV-1 activity.
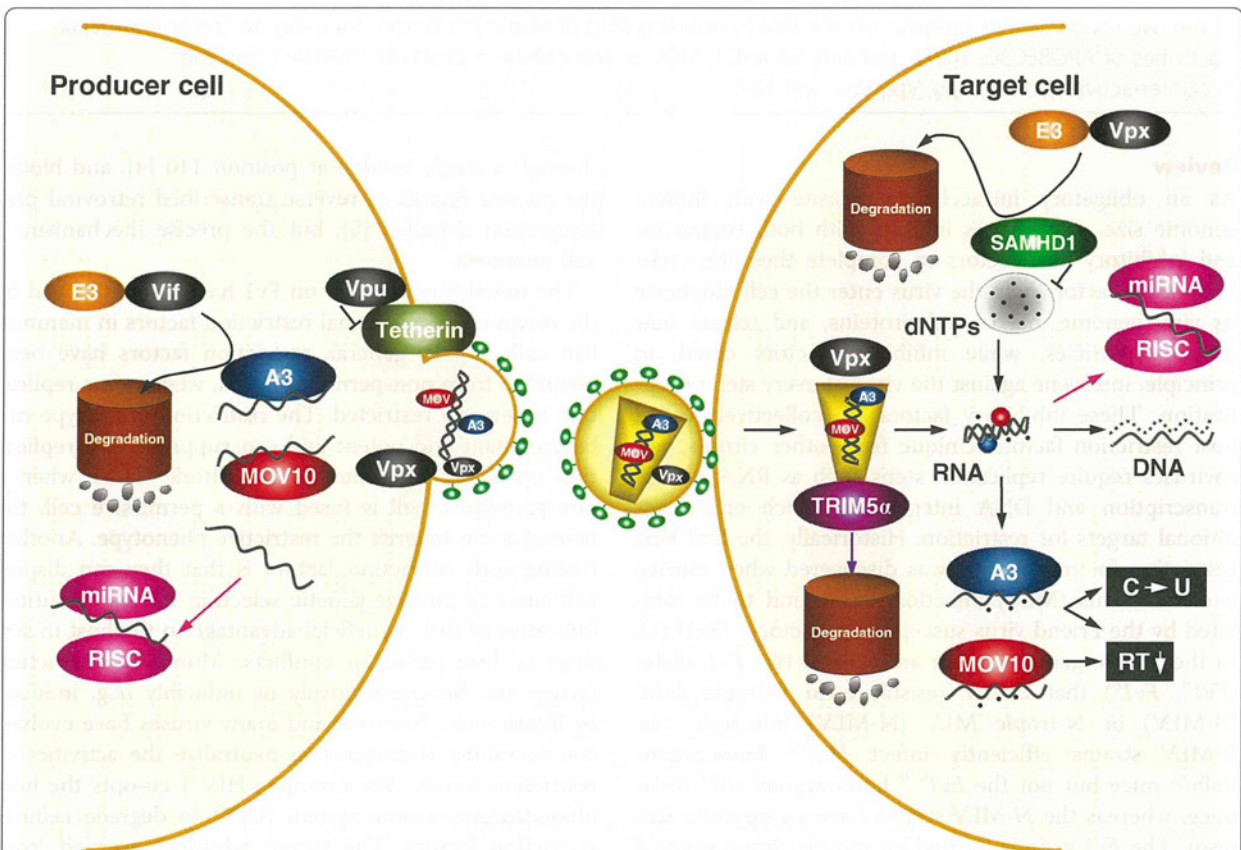
Figure 1 Schematic illustration of the action of host restriction factors during primate lentivirus replication. In viral producer cells, both A3 and MOV10 proteins are packaged into virions via interaction with Gag and RNA. A3 proteins can be targeted by Vif for proteasomal degradation, and viral RNAs can be targeted by specific microRNAs (miRNAs) for suppression. Vpx is also packaged into virion via direct interaction with Gag. In addition, viral release can be inhibited by the cell surface protein tetherin, but it is counteracted by Vpu or Nef (not shown). In target cells, TRIM5α interacts with incoming Gag proteins and triggers premature viral uncoating, resulting in inhibition of viral reverse transcription and nuclear import. Reverse transcription can also be directly inhibited by MOV10 and A3 proteins, or indirectly by SAMHD1 after depleting intracellular dNTP pool. However, SAMHD1 can be neutralized by Vpx through proteasomal degradation. In addition, A3 proteins catalyze C-to-U cytidine deamination on newly synthesized viral cDNA, and viral RNAs can be targeted by miRNAs, which also result in inhibition of viral replication.

Figure 2 Schematic illustration of seven human A3 proteins. Numbers indicate amino acid positions. The zinc-coordinating cytidine deaminase (Z) domain motifs are indicated. The enzymatically active Z domain is presented in dark blue, and the inactive Z domain is in grey. The Vif-interacting motif (red), virion packaging motif (green), and critical residues affecting the enzymatic activity (blue) are all indicated.
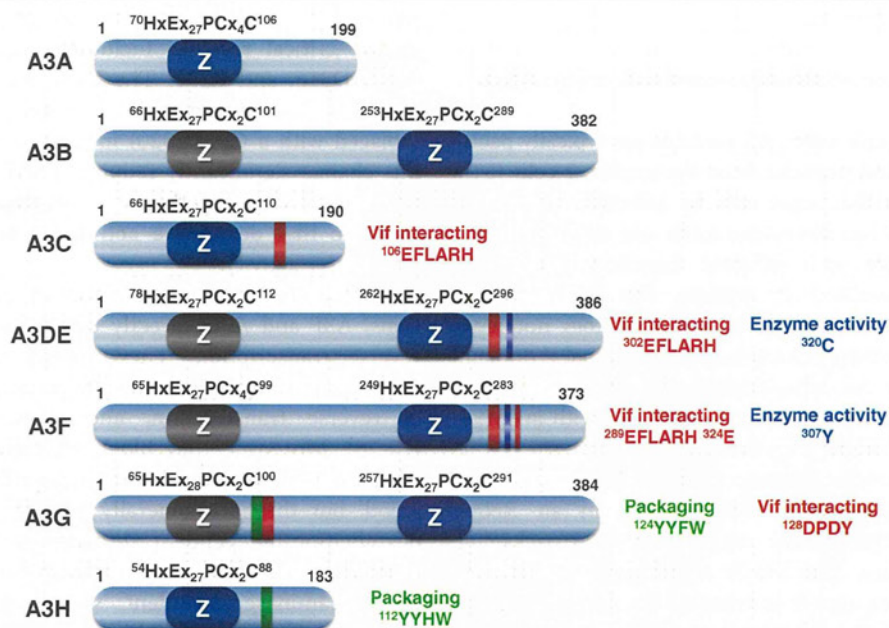
Discovery of A3 protein antiretroviral activity

The discovery of A3's antiviral role originated from the characterization of one HIV-1 accessory gene function. The *vif* gene encodes a ~23 kDa "virion infectivity factor", which is expressed in all lentiviruses except for equine infectious anemia virus (EIAV). Vif is absolutely required for HIV-1 replication in non-permissive cells, although it is not required in permissive cells [12,13]. Vif must be present in the viral producer cells; otherwise, viral replication is strongly blocked at the reverse transcription step in target cells [13]. It was hypothesized that permissive cells either express a Vif-like positive factor, which would replace the *vif* gene function, or that non-permissive cells express an inhibitory factor, which would require Vif for counteraction [14]. A genetic complementation assay performed by fusing the non-permissive cells with permissive cells demonstrated that the non-permissive phenotype is inheritable, indicating that the non-permissive cells express a dominant inhibitory factor [15,16]. By subtraction of cDNAs of non-permissive cells from those of permissive cells, this inhibitor was identified as A3G [17].

The discovery of A3G's anti-HIV-1 activity has invited further investigation on its sister proteins. A3F was identified as the second restriction factor from this family [18,19], and the inhibitory activities of A3B, A3DE, and A3H were subsequently uncovered [20-23]. Among these five anti-HIV A3 proteins, A3B is expressed very poorly in primary tissues due to unexplained reason

[24], and the *A3H* gene is highly polymorphic [25-27]. So far, seven A3H haplotypes (I-VII) have been identified in human populations [26,28]. However, only haplotypes II, V, and VII could be stably expressed, and their total allelic frequencies are less than 30%. In addition, due to a Tyr-to-Cys substitution at position 320, the human *A3DE* anti-HIV-1 activity is reduced by approximately 20-fold [29], although the potent A3DE antiviral activity is still retained in chimpanzees [30]. The endogenous A3DE, A3F, and A3G have the most relevant anti-HIV-1 activity *in vivo* [31], although A3DE and A3F may have weaker activity than A3G [32].

The antiviral activity of A3 proteins extends to other retroviruses such as hepatitis B virus (HBV) [33], SIV [34], HTLV-1 [35], foamy virus [36-38], EIAV [39], MLV [40], and mouse mammary tumor virus (MMTV) [41]. While A3A and A3C do not have anti-HIV-1 activity, A3A strongly inhibits adeno-associated virus (AAV) replication [42], and A3C inhibits SIV replication [43]. In addition, all the A3 proteins inhibit the replication of retrotransposons, although the levels of this activity may vary [44]. *A3* expression is highly inducible by type I interferons (IFNs), particularly in myeloid-derived cells [45,46]. Both *A3DE* and *A3G* genes have been subjected to strong positive selection during evolution [30,47]. In addition, A3 knockout mice are more susceptible to MMTV, MLV, and mAIDS virus infection [41,48-50]. Thus, *A3* genes play an important role in defending

in vivo against both exogenous retrovirus and endogenous retroelements.

Action of A3: deamination-dependent and -independent inhibition

To inhibit viral replication, A3 proteins are typically packaged into retroviral particles from the producer cells [51], and delivered to the target cells by infection (Figure 1). The human A3G has 384 amino acids, and contains duplicative Z domains with different functions (Figure 2). Unlike the C-terminal Z domain, the N-terminal Z domain is a pseudocatalytic domain that does not have any enzymatic activity. However, the N-terminal Z domain has high affinity for RNA-binding and determines A3G packaging into virions [52]. In conjunction with an adjacent ^{124}YYxW^{127} motif [53], the pseudocatalytic site of the N-terminal Z domain interacts with the HIV-1 Gag protein in the nucleocapsid (NC) region, in an RNA-dependent manner [54-61], resulting in A3G packaging into HIV-1 virions. The YYxW motif is also present in other A3 proteins, and it is required for A3H packaging [28], indicating its important role in regulating this critical step of A3 antiviral activity.

After being delivered to target cells, the C-terminal enzymatically active Z domain inhibits viral replication by either cytidine deamination-dependent or -independent mechanisms at the reverse transcription step. The active Z domain can directly deaminate cytosines (C) to form uracils on newly synthesized minus-strand viral cDNAs, leading to changes in the viral sequences [39,62-64]. Since these edited cDNAs contain uracils that are usually not present in DNA molecules, they may be recognized by DNA repair enzymes for degradation. Otherwise, they are copied into the plus strand DNA during DNA synthesis, generating DNA molecules containing G-to-A hypermutations that compromise viral genome integrity. This process is generally termed the cytidine deamination-dependent mechanism. Not every cytosine is mutated, and different A3 proteins have their own dinucleotide preference sequence. For example, A3G prefers to mutate minus strand 5'-CC-to-CU and cause plus strand 5'-GG-to-AG mutation [65]; A3B, A3DE, A3F, and A3H mutate 5'-TC-to-TU and cause 5'-GA-to-AA mutation [22-24,65]; and A3DE also mutates 5'-CG-to-UG and causes 5'-GC-to-AC mutation [22]. All these three types of G-to-A hypermutations are detectable from HIV-1 patients, and it is still unclear which enzyme takes the major responsibility for the 5'-GT-to-AT mutation [66]. The three-dimensional structure of the A3G C-terminal domain has been solved by nuclear magnetic resonance (NMR) and X-ray crystallography [67-69]. This part of the protein shows a globular structure that is formed by five β-sheets and six α-helices and contains a substrate-binding surface. The

X-ray structure further shows a deep groove in this region for substrate binding, which is composed of several critical residues including N244, W285, R313, Y315, D316, and R374. The Y315 residue (or Y307 in A3F) is crucial for deaminase activity. This residue is replaced with a Cys (C320) in human A3DE (Figure 2); this change significantly reduces A3DE anti-HIV-1 activity [29]. Taken together, the findings provide strong evidence that deaminase activity is important to A3's anti-HIV-1 activity.

In addition to the introduction of catastrophic mutations, A3F and A3G directly block the process of viral reverse transcription. They reduce the efficiency of tRNA^{lys3} priming to the viral RNA template, elongation of reverse transcription, and DNA strand transfer [70-76]. Moreover, they block viral integration [75,77]. All these inhibitory mechanisms are still not fully understood, but they are generally thought to arise from a deamination-independent antiviral mechanism. Notably, there is a great deal of confusion in the literature regarding deaminase-dependent versus deamination-independent mechanisms. Although initial investigations suggested that the deaminase activity was not required for antiviral activity, later studies demonstrated that it is required when the A3 proteins are expressed at physiological levels [78-81]. In contrast, it has been consistently observed that A3 proteins block viral replication even in the absence of cytidine deamination, particularly when HTLV-1, AAV, HBV, and retrotransposons are targeted by A3 proteins [33,35,42,82-87]. Thus, the deaminase activity is always required, and in various settings the mechanism is deamination-independent.

Although a different opinion exists [88], several lines of evidence indicate that A3 proteins may require other cellular cofactors. First, APOBEC1, which is the founding member of this family, requires a cofactor. APOBEC1 introduces a premature stop codon in apolipoprotein (apo) B100 mRNA through C-to-U editing to produce a truncated form of this protein [10]. This process requires an interaction with the APOBEC1 complementation factor (ACF) to form a holoenzyme, or so-called editosome to edit the target sequence [89,90]. Thus, A3 proteins may also require a cellular cofactor for target recognition. Second, both A3F and A3G form two distinctive protein complexes: a high molecular mass (HMM) complex over 700 kDa and a low molecular mass (LMM) complex below 100 kDa [91]. The A3G HMM complex contains cellular RNAs and predominantly exists in immortalized cell lines; this complex changes into an enzymatically active LMM complex when treated with RNases. Over a hundred A3G-binding proteins, most of which are RNA-binding proteins, have been identified from these complexes [92-94]. It is conceivable that some of them may

facilitate A3G virion packaging and/or antiviral activity. Third, the cellular expression levels of A3G do not always correlate with its antiviral activity. This phenomenon is particularly remarkable in the human CD4$^+$ CEM-derived T cells. The parental CEM cell line is non-permissive for Vif-deficient (ΔVif) HIV-1 due to A3G expression, and its derivative CEM-SS cells are permissive because they do not express A3G. Notably, its derivative A3.01 and CEM-T4 cell lines express significant levels of A3G, but they can support ΔVif HIV-1 replication [79]. Absence of a critical cofactor may explain why A3 proteins are not active in these cells.
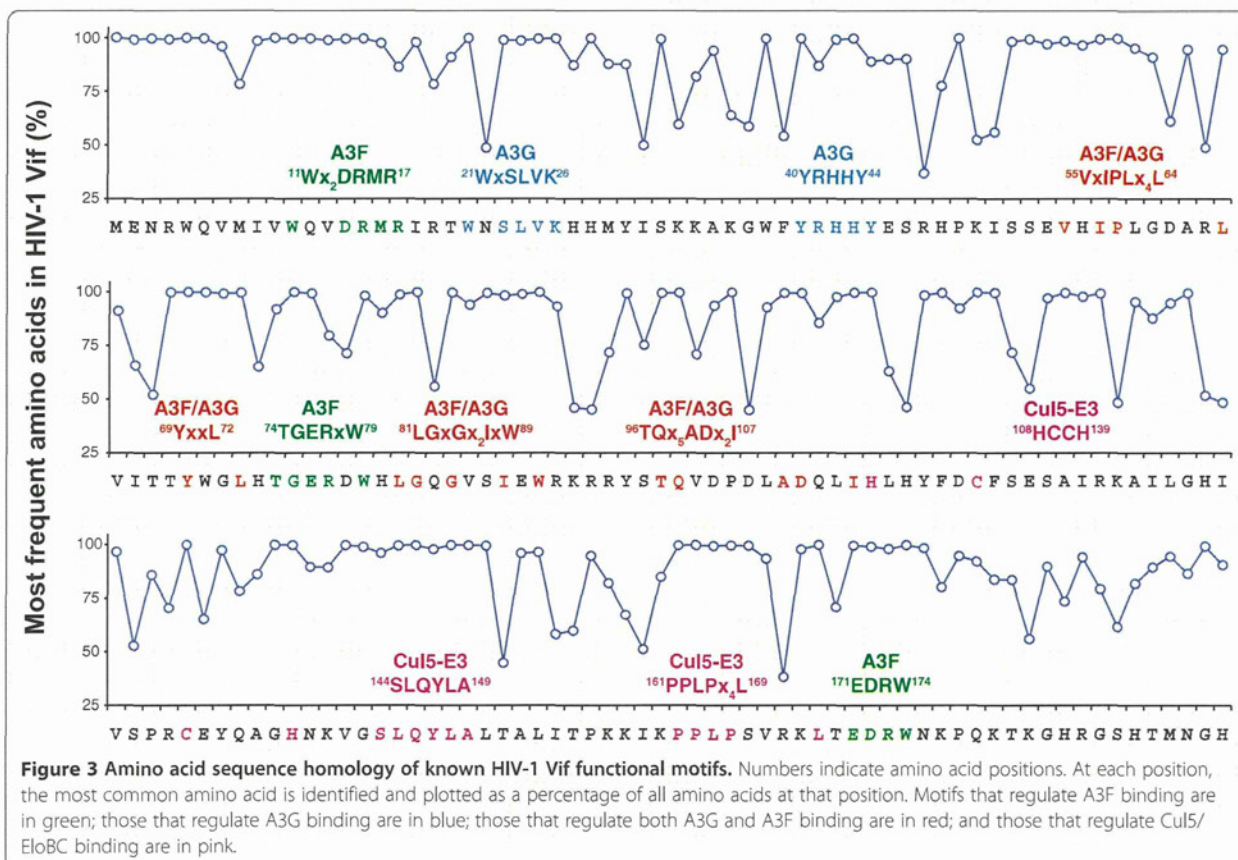
Action of Vif: degradation-dependent and -independent inactivation

Because virion packaging is required for A3G antiviral activity, a critical action of Vif is to exclude A3 proteins from virions in order to protect viral replication [95]. This is achieved by degradation-dependent and/or independent mechanism in the viral producer cells.

The degradation-dependent mechanism hijacks the cellular proteasomal pathway to degrade A3 proteins to ensure that insufficient A3 proteins are packaged into the virions. HIV-1 Vif has 192 amino acids and contains 12 highly conserved motifs (Figure 3). These motifs form

discontinuous surfaces, so that Vif can interact with A3 and E3 ligase. First, Vif interacts with the Cul5 E3 ubiquitin ligase complex, which includes Cullin 5 (Cul5), Elongin B (EloB), and Elongin C (EloC) [96]. This interaction is mediated by its three C-terminal motifs. The ^{108}Hx$_5$Cx$_{17-18}$Cx$_{3-5}$H^{139} motif, which is also called the HCCH zinc finger, binds to Cul5 [97-99]; the ^{144}SLQYLA^{149} motif, which is also called the BC-box, binds to EloC [100,101]; and the ^{161}PPLPx$_4$L^{169} motif, which is also called the Cul box, binds to Cul5 [101,102]. The ^{161}PPLP^{164} subdomain has multiple activities, which not only determine Vif dimerization [103], but also regulate Vif binding to A3G [104-106] and EloB [107]. Second, Vif interacts with its target A3DE, A3F, and A3G via other widely distributed motifs. The ^{21}WxSLVK^{26} [108,109] and ^{40}YRHHY^{44} [110] motifs regulate Vif binding to A3G; the ^{11}Wx$_2$DRMR^{17} [110], ^{74}TGERxW^{79} [111], and ^{171}EDRW^{174} [112] motifs regulate Vif binding to A3F; and ^{55}VxIPLx$_4$L^{64} [111], ^{69}YxxL^{72} [113], ^{81}LGxGx$_2$IxW^{89} [112], and ^{96}TQx$_5$ADx$_2$I^{107} [114] motifs regulate Vif binding to both A3G and A3F. Those motifs that regulate Vif interaction with A3F also regulate Vif interaction with A3DE [115].

The Vif-interactive domains on A3DE, A3F and A3G have also been identified. Although A3F and A3G share 50% amino acid sequence identity, they use different

Figure 3 Amino acid sequence homology of known HIV-1 Vif functional motifs. Numbers indicate amino acid positions. At each position, the most common amino acid is identified and plotted as a percentage of all amino acids at that position. Motifs that regulate A3F binding are in green; those that regulate A3G binding are in blue; those that regulate both A3G and A3F binding are in red; and those that regulate Cul5/EloBC binding are in pink.

domains to interact with HIV-1 Vif. Vif binds to the A3G N-terminal subdomain ^{128}DPD^{130} [53], and it recognizes the A3F C-terminal ^{289}EFLARH^{294} domain and a residue ^{324}E [116-118]. The EFLARH domain, also present in A3C and A3DE, further determines their Vif sensitivity (Figure 2) [118]. These interactions assemble the A3-Vif-Cul5 E3 protein complex, which induces polyubiquitylation and proteasomal degradation of these A3 proteins.

The A3G polyubiquitylation sites have been mapped to lysine residues 297, 301, 303, and 334 [119]. However, there is evidence that A3G is polyubiquitylated through the N-terminus [120,121]. Notably, Vif itself is polyubiquitylated by the same E3 ligase [100] and is degraded via the proteasome [122-124]; and Vif polyubiquitylation is very critical for A3G proteasomal degradation [125]. To initiate proteasomal degradation, a protein should have at least two signals: an attached polyubiquitin chain and an unstructured region (USR) [126]. The polyubiquitin chain is required for proteasome recognition, and the USR allows the protein to enter the narrow proteasome entrance channel to initiate degradation. Although both signals can be present in the same protein, they can also work in *trans* in two different proteins that bind to each other. A protein missing any signal may remain stable, until it binds to an adaptor protein to compensate for the missing signal. Since A3G is well structured, it is possible that A3G degradation is dependent on the polyubiquitin chain on Vif, but not on A3G itself. Thus, more investigation is needed to understand better whether A3G polyubiquitylation is critical for its neutralization, and how Vif polyubiquitylation contributes to this process.

The action of Vif is highly species-specific. Vif from HIV-1 only inactivates A3G from human; and Vif from SIV, isolated from African green monkey (AGM), does not inactivate human A3G. Nevertheless, Vif from SIV isolated from rhesus macaque (MAC) inactivates A3G from all humans, African Green monkeys, and macaques [127]. The resistance of agmA3G to HIV-1 Vif is due to a single mutation in the ^{128}DPD^{130} motif of A3G, which is adjacent to A3G ^{124}YYxW^{127} packaging motif [53]; notably, this motif is changed to ^{128}KPD^{130} in the agmA3G [128-130]. In addition, a N-terminal domain in HIV-1 Vif, ^{14}DRMR^{17}, determines Vif activity for A3G from different species [131].

Recently, the core-binding factor β (CBF-β) was identified as a critical cofactor for Vif [132,133]. CBF-β forms a heterodimer with the RUNX transcription factors and increases the complex binding to the target DNA [134]. Knockdown of CBF-β expression in 293T cells was found to compromise Vif's ability to trigger A3G degradation, but it is still controversial whether CBF-β stabilizes Vif protein itself, or simply induces Vif

conformational changes to stabilize the Vif-Cul5 complex. In addition, its critical role in Vif function needs to be verified in human T cells.

Many extant observations can be explained by the A3G degradation-dependent inhibition model; however, emerging evidence suggests that this is not the only mechanism for Vif to neutralize A3G. For example, if Vif-induced A3G degradation is fully responsible for A3G inactivation, then the levels of A3G reduction by Vif in viral producer cells should be proportional to those in virions. In fact, the levels of A3G reduction by Vif in virions are always much more pronounced than those in viral producer cells. Thus, Vif is apparently able to block A3G encapsidation in the absence of an induction of degradation [135]. The existence of this degradation-independent mechanism is supported by two mutational studies. The A3G C97A mutant is resistant to Vif induced proteasomal degradation, but its activity is still neutralized by Vif [136]. In addition, Vif proteins that show different ability to degrade A3G exhibit similar efficiencies in neutralizing A3G [137]. In fact, Vif can inactivate A3G enzymatic activity in *E. coli*, which does not have an UPS for protein degradation [138]. Thus, it seems reasonable that Vif employs both degradation-dependent and -independent mechanisms to counteract A3G's antiviral activity.

A3 and HIV evolution

Because HIV-1 replicates in A3-expressing cells in the presence of Vif, it has been thought that Vif completely counteracts A3's antiviral activity. However, the virus may not benefit from a complete inhibition of A3 activity. Although lethal mutations inhibit viral replication, sublethal mutations can promote viral propagation in hosts through a promotion of genome evolution. HIV-1 is notorious for its rapid evolution. It maintains an optimal mutation rate that allows viral escape from adaptive immunity and development of drug resistance. Among various mutations, the appearance of G-to-A hypermutations is the most frequent [139], which may explain why the HIV-1 genome is extremely A-rich [140,141]. Although G-to-A hypermutations have been attributed to the errors of the low-fidelity viral reverse transcriptase and viral genomic recombination, it has been recently appreciated that A3 proteins are significantly responsible for this type of mutation. Analyses of viral hypermutations at population levels have detected both GG→AG or GA→AA dinucleotide motifs for introducing hypermutations [142-145]. In addition to inhibition of viral replication, A3 proteins may create beneficial hypermutations, which generate viral quasispecies and diversify viral genomes [146-148]. Both A3G and A3F facilitate viral adaptation to the new drug treatment environment if their activities are not completely neutralized by Vif

[149,150]. Several mechanisms may cause A3 activity to not be completely neutralized by Vif. First, the relative levels of protein expression may determine neutralization. Because A3 expression is interferon inducible, A3 expression level may well exceed the Vif expression level at the early stage of viral replication, which may make it impossible for Vif to neutralize fully A3 proteins. Second, the degree of neutralization may be regulated at the levels of protein quality. The *vif* gene itself could become the target for hypermutation and become polymorphic during natural infection. Any mutation that interferes with Vif binding to A3 and/or Cul5 E3 ligase will compromise A3 neutralization by Vif, resulting in incomplete A3 inactivation [151,152]. Thus, it is anticipated that a balance between Vif and A3 proteins has to be established over time to maintain optimal viral fitness and composition of viral quasispecies *in vivo*. It is also conceivable that viral replication could be inhibited if this balance is disrupted by pharmacological intervention [153], either by a complete inhibition of Vif activity [154-156], or by a commensurate inhibition of A3 enzymatic activity [157].

TRIM5

TRIM5 belongs to the tripartite motif (TRIM)-containing protein family, which has over 100 members [158]. All TRIM proteins have three motifs, including a N-terminal RING finger motif, followed by one or two B-box motifs, and then by a coiled-coil motif [159-161]. The RING, B-box, and coiled-coil motifs are also collectively called the RBCC domain. The C-terminal regions of these proteins vary, but most of them have a SPRY (also known as B30.2) motif. The RING finger binds to two zinc atoms, and usually has E3 ubiquitin ligase activity; the B-box and CC domains promote protein oligomerization. Human TRIM5 has six major isoforms, with the α isoform most abundantly expressed (~50%) [162]. Human TRIM5α has 493 amino acids, and it is the only isoform that has the C-terminal SPRY domain (Figure 4). Its RING and B-box 2 domains are separated by the linker 1 (L1) region, and its coiled-coil and SPRY domains are separated by the linker-2 (L2) region. The SPRY domain has been functionally replaced with another host protein cyclophilin A (CypA) in a number of monkey

species, resulting in another protein TRIM-CypA [163]. TRIM5α and TRIM-Cyp are the only isoforms that have antiretroviral activity and inhibit retroviral replication in a species-specific manner.

Discovery of TRIM5 antiretroviral activity

Although Fv1 is only expressed in mice, the N-MLV strains encounter an Fv1-like restriction in non-murine species including humans, and this unknown MLV inhibitor was named restriction factor 1 (Ref-1) [164]. Ref1 also inhibits EIAV replication in human cells [165]. HIV-1 and some SIV strains encounter another Fv1-like restriction when they infect some non-human species. For example, HIV-1 replication is inhibited in the Old World Monkeys (rhesus macaques, African green monkeys) and New World Monkeys (squirrel monkeys, common marmosets); SIV (SIVmac) infection is blocked in the squirrel monkeys. The unknown HIV-1 and SIV inhibitors were named lentivirus susceptibility factor 1 (Lv1) [166,167]. Fv-1, Ref-1, and Lv-1 share remarkable similarities in their viral inhibition. First, they all target an early post-entry step. Both Ref-1 and Lv-1 act at steps before or after reverse transcription, whereas Fv1 acts at a step after reverse transcription but before integration. Second, they all target viral CA proteins. The same single CA 110 residue that differentiates between N- and B-tropism in mice also determines MLV tropism in human cells. Similarly, HIV-1 and SIVmac restriction in some primate cell lines is determined by sequences within the CA-p2 region of Gag [167-169]. Third, all these restrictions can be released by a high multiplicity of infection (m.o.i.), indicating that they are saturable. In due course, the Lv1 restriction activity was first identified as the TRIM5α protein from rhesus monkey cells and later as the TRIM-Cyp fusion from the owl monkey cells [170,171]. Subsequently, Ref-1 was identified as the human TRIM5α protein [172]. Thus, Ref-1 and Lv1 are specifies-specific TRIM5α proteins that have different activities against different retroviruses.

TRIM5 E3 ubiquitin ligase activity

The TRIM5 RING finger motif features a cysteine-rich consensus that contains two interleaved Zinc-binding

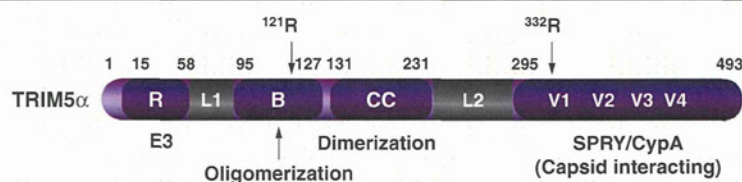**Figure 4 Schematic illustration of human TRIM5α protein.** Numbers indicate amino acid positions. The RING finger (R), B-box (B), coiled-coil motif (CC), two linkers (L1, L2), and SPRY domain are indicated. Four variable regions in SPRY, the critical residue (R121) in the B domain that determines oligomerization, and the critical residue (R332) in V1 region that determines species-specific Gag binding are also indicated.

sites. This motif may simultaneously bind ubiquitination enzymes and their substrates, and hence exhibits an E3 ligase activity. For example, the RING-box-1 (Rbx1) is an essential component of the Skp1-cullin1-F box (SCF) complex, which is a multi-protein E3 ubiquitin ligase that regulates cell cycle [173]. Human TRIM5 has a ββα RING fold with a putative E2-binding region [174]. The TRIM5 E3 ligase activity was first demonstrated in TRIM5σ [175], and recent studies suggest that this activity may contribute to the restriction activity (see below) [176,177].

TRIM5α triggers self-polyubiquitylation using UbcH5 as an E2-conjugating enzyme [178], but the role of this self-polyubiquitylation is still not completely clear. TRIM5α is relatively unstable in cells, with a protein half-life only about 50 to 60 minutes; but current data suggest that the proteasomal pathway is not responsible for its rapid turnover [178] and that this rapid turnover is not required for TRIM5α's antiretroviral activity [179]. Nevertheless, the TRIM5α turnover is further enhanced when cells are infected with restriction-sensitive viruses. In this setting, the proteasomal pathway is responsible for its enhanced degradation [180] and TRIM5α has been found to be associated with proteasomal subunits in cells [181]. Interestingly, the inhibition of the proteasomal pathway does not significantly disrupt overall TRIM5α restriction, despite disrupting TRIM5α-mediated inhibition of viral reverse transcription [182]. Accordingly, two independent inhibitory mechanisms have been proposed. The first is accelerated viral uncoating by TRIM5α interaction with CA, which triggers premature uncoating and proteasomal degradation of the viral reverse transcription complex. The second is a block to the nuclear translocation of the viral preintegration complexes, which has also been described for Fv1 [182].

In addition to a direct inhibition of viral replication, TRIM5α can be a signaling molecule that activates the NF-κB pathway [183]; this pathway is shared in several receptor-signaling routes including those used by Toll-like receptors (TLRs). Stimulation of TLRs induces autophosphorylaton and activation of TGF-activated kinase 1 (TAK1), and TAK1 in turn activates IκB kinase, leading to the activation of NF-κB. TAK1 can be directly activated by unanchored (free C-terminus) K63 polyubiquitin chains [184]. In fact, activated TRIM5α interacts with UBC13 and UEV1A E2 enzymes and triggers the production of unanchored K63-linked ubiquitin chain, resulting in the activation of TAK1 [176], and this activity also extends to several other TRIM family members [185]. In addition, TRIM5α was recently reported to traffic through the nucleus [186], which may be related to this activity. Thus, by interacting with viral cores, TRIM5α induces TAK1 autophosphorylation and activates

the NF-κB pathway, which indirectly defends against viral infection via signaling through the innate pattern-recognition receptor-mediated immune responses.

The antiviral activity of TRIM5α can be disrupted by treatment with As_2O_3 (arsenic trioxide), but the mechanism for this effect is still unclear [187,188]. Trivalent arsenic (As^{III}) has very high affinity for free thiols, which are present in the Cys residues in TRIM5α. As^{III} has very broad biological activities, which are largely mediated by direct interaction with Cysteines in target proteins [189]. When these Cys residues are located in close proximity, this interaction results in their cross-linking, which causes protein conformational changes through S-As-S bond formation. As^{III}-induced conformational changes have different impacts on protein function. Notably, As^{III} directly binds to the Cys residues in the RING motif of TRIM19, which is also known as promyelocytic leukemia (PML) protein, and this interaction makes PML more accessible to SUMOylation, resulting in enhanced PML degradation [190]. It remains to be determined whether As^{III} disrupts TRIM5α's antiviral activity by a similar mechanism.

TRIM5 oligomerization

Unlike the RING domain, the B-box 2, coiled-coil, and SPRY domains are absolutely required for TRIM5α's antiretroviral activity [191,192]. TRIM proteins share a common feature in supporting the structure of various cytoplasmic and nuclear bodies through self-association [160]. TRIM5α proteins are found in cytoplasmic bodies, although these cytoplasmic structures do not directly contribute to antiviral activity. TRIM5α oligomerizes at two different levels, which are determined by the B-box 2 and coiled-coil domains. The coiled-coil domain determines TRIM5α dimerization, which is a lower-order oligomerization [193,194]; the B-box 2 domain, particularly the residue R121, triggers a higher-order oligomerization through association of these dimers (Figure 4) [195]. Dimerization is essential for higher-order oligomerization, and the RING domain and linker 2 region are also involved in the oligomerization process [196]. HIV-1 has a cone-shaped viral core, which is supported by CA proteins that form pentagonal and hexagonal lattices, where the CA N-terminal domain (NTD) forms either hexameric or pentameric rings, and the C-terminal domain (CTD) forms symmetric homodimers that connect the rings into lattices [197]. TRIM5 proteins spontaneously assemble into hexagonal lattices, which match the symmetry of CA lattices, and this assembly can be enhanced by recombinant CA proteins that have already preformed the conical structure [198,199]. Thus, the higher-order TRIM5α structure increases its binding for the viral CA protein.

As introduced earlier, TRIM5α accelerates viral uncoating to block viral replication. Uncoating occurs within several hours after viral entry, which releases the viral genome from the viral core by removing capsid and envelope [200]. After HIV-1 entry, the CA proteins are detectable in the cytosolic fraction as both pelletable and soluble forms. The pelletable form could be derived from the intact cores that have not been uncoated, and the soluble form could be derived from uncoated cores. TRIM5α accelerates the conversion of the viral CA proteins from pelletable to soluble forms [201]. Moreover, incubation of the preassembled CA recombinant proteins with TRIM5α resulted in the disruption of the CA conical structure, likely by weakening the CA CTD-CTD interfaces between hexamers [202]. Thus, it is suggested that TRIM5α could form the hexagonal structures on top of the capsid lattices, which disrupts the core and then further destruction of viral proteins could occur via the proteasomal machinery. In this model, a linkage exists between TRIM5α degradation by the proteasome and accelerated viral uncoating by TRIM5α. Since TRIM5α is associated with the proteasomal machinery [181] and its contribution to uncoating is blocked by proteasomal inhibitors, proteasomes should be engaged in TRIM5α enhancement of core disassembly.

TRIM5 cross-species activity

Unlike other restriction factors, TRIM5α normally does not inhibit retroviruses isolated from the same host species. For example, human TRIM5α (hsTRIM5α) has a very weak activity for HIV-1, but it strongly inhibits EIAV and N-MLV; TRIM5α from rhesus monkeys (rhTRIM5α) does not inhibit the SIVmac strains, but it strongly inhibits HIV-1 and some other SIV strains [171]. However, a single residue exchange in the hsTRIM5α SPRY domain (R332) with the corresponding rhTRIM5α residue (P334) is sufficient to lead to HIV-1 restriction by the altered hsTRIM5α [203,204]. This residue is among a cluster of residues, which are located in the first variable region (V1) and have been found to be under strong positive selection (Figure 4) [205]. Analysis of TRIM5α sequences from different species identified four variable regions (V1, V2, V3, V4) in the SPRY domain [203,206], and three of them (V1, V2, V3) contribute to TRIM5α specificity of retrovirus restriction [191]. The rhesus macaque *TRIM5* gene is particularly polymorphic in the SPRY region. A three-residue replacement with another residue in the V1 region results in TFP/Q polymorphisms, and in some cases, the entire SPRY region is replaced with the *CypA* gene to produce a new TRIM-CypA protein. Accordingly, the *TRIM5* gene is classified into three allelic groups based on the SPRY domain in the various species: $TRIM5^{CypA}$, $TRIM5^{TFP}$, and $TRIM5^Q$ [207]. These polymorphisms have significant impact on SIV cross-species infection

[207-209]. Consistent with host specificity, a large population study has found that common human variants of TRIM5α has little to no effect on HIV-1 disease progression [210], suggesting that the role of this protein in the human genome is not to mitigate HIV-1 infection and instead serves a not yet understood function.

Although the SPRY domain is not required for hexagonal array formation, compelling evidence suggests that this domain is directly engaged in the interaction with CA proteins; and this interaction specifies cross-species restriction activity [188,201]. Detection of direct TRIM5α-CA interaction has been difficult, because TRIM5α does not bind to monomeric or soluble CA proteins [211]. This interaction only becomes detectable in a SPRY-dependent manner when TRIM5α is directly incubated with purified viral cores or pre-assembled recombinant CA proteins [188,201]. These results suggest that TRIM5α recognizes CA proteins in a conformation-dependent manner, which is consistent with a previous finding that only stable and mature cores could neutralize the TRIM5α antiviral activity [212]. A structural analysis confirmed the direct interaction between the SPRV domain and CA protein, and that the variable regions, particularly the V1 region, are responsible for this interaction [213]. This study also uncovered that a single-site SPRY-CA binding is weak, and that optimal interaction involves multiple CA epitopes, which may explain why multivalent binding within the spacing of CA lattice is required for viral restriction.

Tetherin

Tetherin was originally identified (and termed HM1.24, BST-2, or CD317) as a specific cell surface marker of terminally differentiated B-cells, through screening of mouse monoclonal antibodies raised against the human plasma cells [214]. The gene encoding this protein was independently cloned from the human rheumatoid arthritis-derived synovial cells and termed BST-2 [215]. The protein was proposed to potentially serve as a target antigen for the immunotherapy of multiple myeloma since its mouse and humanized monoclonal antibodies showed anti-tumor activity with antibody-dependent cellular cytotoxicity both *in vitro* and *in vivo* [216,217]. The protein, re-designated as CD317, was then found to be highly expressed in B cells at all differentiation stages, and in bone-marrow CD34$^+$ cells, and in T cells [218].

Tetherin is an interferon (IFN)-inducible type II membrane protein, consisting of a short amino-terminal cytoplasmic tail (CT) followed by an α-helical transmembrane (TM) domain, a coiled-coil extracellular (EC) domain, and a carboxy-terminal glycophosphatidylinositol (GPI) component that acts as a second membrane anchor [219] (Figure 5A and 5B). This double-anchored form determined by the TM and the GPI anchor is

Figure 5 (See legend on next page.)
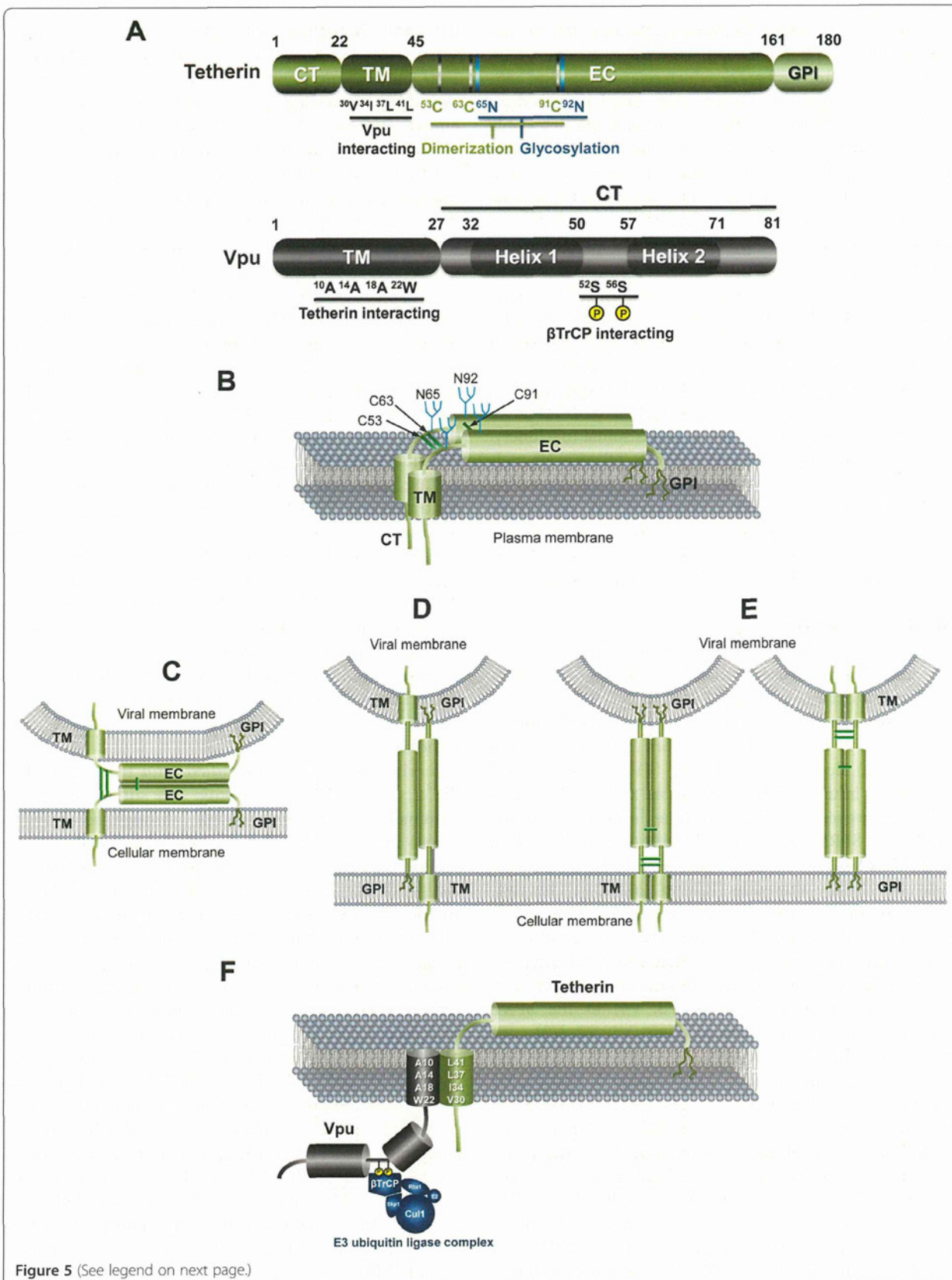
(See figure on previous page.)
Figure 5 Configuration models of human tetherin. (A) Schematic illustrations of tetherin and Vpu. Numbers indicate amino acid positions. Critical residues of each protein are indicated. **(B)** Structure of tetherin. Tetherin comprises a short amino-terminal cytoplasmic tail (CT), followed by an α-helical transmembrane (TM) domain and a coiled-coil extracellular (EC) domain that is linked back to the plasma membrane by a carboxy-terminal glycophosphatidylinositol (GPI) anchor. The EC domain contains N-glycosylation sites and cysteine residues involved in disulfide-bond formation. **(C-F)** Configuration models of tetherin. **(C)** The EC self-interaction model. Individual tetherin monomers are anchored at both ends to the same membrane, with interaction between the ECs of cell-bound and virion-bound monomers. **(D)** Anti-parallel membrane-spanning model. Monomers are anchored in both membranes with opposing orientations. **(E)** Parallel membrane-spanning model. Monomers are anchored in both membranes with the same orientation. **(F)** HIV-1 Vpu and tetherin interact through their TM domains. Key amino acids involved in the interaction are depicted in the TM helices. Interaction of Vpu's CT with the E3 ubiquitin (Ub) ligase via the βTrCP subunit is required for Vpu-induced tetherin down-regulation.

unusual and only shared with a minor isoform of the prion protein [220]. Three cysteine residues located at the EC domain form intermolecular disulfide bonds [221], resulting in homodimerization of the protein [222]. At the cell surface, tetherin is located in cholesterol-rich lipid microdomains, also termed lipid rafts, through the GPI anchor. The TM domain resides outside the lipid rafts [219], placing the CT in a suitable position to indirectly interact with the actin cytoskeleton [223]. Tetherin is physiologically endocytosed from lipid rafts in a clathrin-dependent manner through interaction with α-adaptin of the AP-2 complex [224].

Discovery of tetherin's antiviral activity

Early studies on HIV-1 replication showed that the accessory viral protein U (Vpu) was required for efficient viral particle release in a cell-type dependent manner [225,226]. This led to the hypothesis that the requirement was due to either the existence of unknown restriction factor(s) or the lack of cofactor(s) in non-permissive cells. Some years later, experiments generating heterokaryons between permissive and non-permissive cells provided the answer [227]; the latter cells likely express restriction factor(s) that could be counteracted by Vpu [228], exactly like Vif-responsive cells express A3G that is counteracted by Vif. It was then shown that Vpu-deleted HIV-1 particles captured at the cell surface were detached by treatment with subtilisin protease, implying that the endogenously expressed host restriction factor is a membrane-associated cell-surface protein [229]. Soon afterwards, this endogenous factor was found to be IFN α-inducible and indeed its activity could be overcome by Vpu [230]; cDNA microarray analyses of messenger RNAs in IFN-α-treated and untreated cells finally identified HM1.24/BST-2/CD317 as the restriction factor, which was re-termed "tetherin" because of its direct tethering function at the cell-surface [231]. Subsequently, it was demonstrated that Vpu-induced down-regulation of tetherin from the cell surface explains its counteraction of the antiviral activity of tetherin [232]. The growing list of enveloped viruses restricted by tetherin includes filoviruses, arenaviruses, paramyxoviruses, gamma-herpesviruses, rhabdoviruses, and a wide array of retroviruses from several mammal host species [233-239]. These restrictions occur not only *in vitro* but also in natural target cells *in vivo* [234,240].

Antiviral mechanisms of tetherin

Tetherin efficiently blocks the release of Vpu-defective HIV-1 virions by directly tethering them to the surface membranes of virus producer cells. Captured virions are internalized by endocytosis, and subsequently accumulate into CD63-positive endosomes, and probably are degraded in the lysosomes [230,231]. Structurally, the protein's two membrane anchors formed by the N-terminal TM domain and the C-terminal GPI anchor, together with the conformational flexibility provided by the homodimerized EC domain, are key for the direct tethering mechanism required for the above process. In fact, a totally artificial tetherin-like protein consisting of structurally similar domains from three unrelated heterologous proteins (the CT/TM, EC, and GPI anchor from different proteins) reproduced tetherin's antiviral activity by inhibiting the release of Vpu-deleted HIV-1 and Ebola virus-like-particles, despite their lack of sequence homology with native tetherin [241], suggesting that the configuration of tetherin at the cell surface, but not its primary sequence, is important for the antiviral activity. Lipid raft localization of tetherin, which is determined by the GPI anchor, is in accordance with the preferential site for budding of enveloped viruses [242,243]. Indeed, tetherin has been reported to be enriched at the virological synapse [244], but its role in cell-to-cell transfer of viruses remains controversial [245-249]. As another function, it has very recently been shown that tetherin acts as a viral sensor for the presence of viral infection, inducing NFκB-dependent proinflammatory gene expression [250], an activity that described for TRIM5α [176].

With regard to the configuration of tetherin, several models have been proposed: (1) The EC self-interaction model (Figure 5C) — individual monomers of tetherin are anchored at both ends (TM and GPI anchor) to the

same cellular or viral membrane, and the EC domains of cell-associated and virion-associated monomers are bound through disulfide bonds; (2) The membrane-spanning model (Figure 5D and E), both ends are anchored into the opposite side of the membranes (cellular and viral); and theoretically, the dimerized monomers of tetherin in this model can be formed in either an anti-parallel or parallel configuration. The EC self-interaction and anti-parallel membrane-spanning models are supported by the experiment in which the cleavage of the GPI anchor by enzymatic treatment with Pi–PLC did not relieve the restricted virions at the cell surface, suggesting that each GPI-anchorless monomer that is dimerized still anchors at the different sides of the plasma membrane trough the TM domain [251]. On the other hand, electron microscopy studies have revealed that the actual distance of the gap between viral and cellular membranes is larger than the estimated size of that in the EC self-interaction model [241,252,253]. This evidence strongly supports the (anti-parallel or parallel) membrane-spanning model. Most importantly, combined analyses of high-resolution crystallography and small-angle X-ray scattering-based modeling finally demonstrated that the structure of tetherin's coiled coil EC domain is indeed a parallel homodimer [252,254-256]. Taken altogether, it is likely that the parallel membrane-spanning configuration model may correspond to the configuration of the antiviral state of tetherin at the cell surface (Figure 5E).

Action of HIV-1 Vpu

Vpu, which is encoded in the genomes of HIV-1 and a few SIV strains, is an 81-amino acid type I transmembrane protein. It comprises an amino-terminal single TM α-helix domain that also acts as an uncleaved signal peptide, and a carboxy-terminal CT domain in which two cytosolic α-helices are separated by a short flexible connector loop. Vpu mediates proteasomal degradation of CD4 by interacting with newly synthesized CD4 molecules in the endoplasmic reticulum, together with the β-transducin repeat-containing protein (βTrCP) 1 and 2 subunits through its phosphoserine residues in the CT domain [257,258] (Figure 5A and 5F). This βTrCP dependency of Vpu is only partially common to the anti-tetherin activity since βTrCP-binding-defective mutant viruses still retain half of the wild-type activity [259]. Thus, it seems likely that some unknown cellular co-factors other than βTrCP proteins might be required for Vpu to inhibit the antiviral activity of the restriction factor [232,260].

The models of intracellular sites of Vpu's action in tetherin down-regulation have been controversial. First, it was proposed that Vpu interferes with the membrane transport of newly synthesized tetherin by sequestering

the restriction factor in the trans-Golgi network (TGN) [261-263]. Second, Vpu might be able to block the recycling of tetherin by sequestering the latter protein in the recycling endosomes after its internalization from the cell surface [261,263,264]. Third, it was suggested that Vpu might directly internalize tetherin from the cell surface leading to lysosomes [259,265,266], possibly in a cell-type-dependent manner [267]. These three models may not be mutually exclusive, but rather it is likely that each antagonistic model of Vpu is operative to counteract tetherin to varying degrees in different cellular contexts.

In terms of the intracellular fate of tetherin, Vpu-induced down-regulation of the restriction factor might be mediated in part through proteasomal degradation [268-270]. This possibility is based on experiments in which the treatment by proteasomal inhibitors resulted in increased levels of tetherin and loss of Vpu-mediated viral release enhancement. However, prolonged incubation with the inhibitors leads to the depletion of the free ubiquitin pool, affecting both proteasomal and lysosomal degradation [271]. Indeed, the latter degradation pathway has been suggested by evidence that the treatment with inhibitors of the lysosomal pathway blocks the Vpu-mediated tetherin degradation [259,264,266], resulting in a clear colocalization of these two proteins to lysosomal compartments [259,265,267]. In accordance with this, it has been reported that tetherin is constitutively degraded in lysosomes by HRS, a key component of the ESCRT-0 complex that sorts ubiquitinated membrane proteins to lysosomes, and this is accelerated by interaction with Vpu [266]. As another explanation, without inducing any degradation, Vpu simply might sequester either *de novo* or recycled tetherin in the TGN and/or the recycling endosomes plus the TGN, respectively [262,263,272,273], as described above.

The ability of Vpu to bind tetherin through TM-TM interaction is crucial for viral antagonism of this restriction factor [259,268,274,275]. This interaction is highly specific at the amino acid level requiring residues I34, L37, and L41 of tetherin [276] and A14, A18 and W22 of Vpu [277] (Figure 5A and 5F) on the hydrophobic faces of the helices that contribute an interactive surfaces. Recent NMR spectroscopy analysis showed that V30 of tetherin and A10 of Vpu (Figure 5A and 5F) together with the aforementioned residues contribute to form an anti-parallel, lipid-embedded helix-helix interface [278]. Importantly, species specificity of tetherin antagonism by primate Vpu proteins is determined by their TM-TM interaction. Indeed, non-human primate tetherin proteins are mostly insensitive to Vpu antagonism [269,279-281], due to the difference of the amino acid positions 30–45 of the TM sequence [268,279,282] that correspond to the interaction surface as described above.